Display
D.

AMERICAN
JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXXIV, Number 3
JULY, 1952

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

CONTENTS

The AMERICAN JOURNAL OF MATHEMATICS will appear four times yearly.

The subscription price of the JOURNAL for the current volume is $7.50 (foreign postage 50 cents); single numbers $2.00.

A few complete sets of the JOURNAL remain on sale.

Papers intended for publication in the JOURNAL may be sent to any of the Editors.

Editorial communications should be sent to Professor AUREL WINTNER at The Johns Hopkins University.

Subscriptions to the JOURNAL and all business communications should be sent to THE JOHNS HOPKINS PRESS, BALTIMORE 18, MARYLAND, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

GE
31
47
55
78
87
97
36
45
56
57
76
79
33
37
01
23
26
37

gn

rs.
ns

to

cial

ON THE COHOMOLOGY THEORY FOR ASSOCIATIVE ALGEBRAS.* [1]

By I. H. Rose.

1. Introduction. Cohomology theory in general (cf. [1]) concerns itself with the following situation, which arose originally in topology:

$C^0, C^1, \cdots, C^n, \cdots$ is a sequence of abelian groups, $\delta^0, \delta^1, \cdots, \delta^n, \cdots$ a sequence of homomorphisms such that δ^n maps C^n into C^{n+1} and such that $\delta^{n+1}\delta^n = 0$.

In this situation the elements of C^n are called n-dimensional cochains; the kernel of δ^n is denoted Z^n and its elements are called n-dimensional cocycles; the image of δ^n is denoted B^{n+1} and its elements are called $(n+1)$-dimensional coboundaries; B^0 is defined to be the zero element of C^0.

Since $\delta^{n+1}\delta^n = 0$, it follows that $B^n \subset Z^n$. We may therefore define the "cohomology" group $H^n = Z^n/B^n$; two n-dimensional cocycles are called cohomologous if their difference is a coboundary. The symbol δ is used to represent any of the homomorphisms δ^n, since this leads to no ambiguity.

The cohomology theory for associative algebras specializes the general situation as follows.

Let F be a field over which are defined an associative algebra A and a vector space P. Suppose further that P is a two-sided A-module, i. e. for each $a \,\varepsilon\, A$, $p \,\varepsilon\, P$ there are defined elements $a \cdot p$, $p \cdot a \,\varepsilon\, P$ which are bilinear functions of a and p such that

$$a_1 \cdot (a_2 \cdot p) = a_1 a_2 \cdot p, \qquad (p \cdot a_1) \cdot a_2 = p \cdot a_1 a_2, \qquad a_1 \cdot (p \cdot a_2) = (a_1 \cdot p) \cdot a_2.$$

Now for $n > 0$ define $C^n = C^n(A, P)$ as the vector space of all n-linear functions of n variables mapping A into P, and define $C^0 = C^0(A, P) = P$. Furthermore, for $n > 0$ and $f \,\varepsilon\, C^n(A, P)$ define δf such that:

$$\begin{aligned}\delta f(a_1, \cdots, a_{n+1}) &= a_1 \cdot f(a_2, \cdots, a_{n+1}) - f(a_1 a_2, a_3, \cdots, a_{n+1}) \\ &+ \sum_{i=2}^{n-1} (-1)^i f(a_1, \cdots, a_{i-1}, a_i a_{i+1}, a_{i+2}, \cdots, a_{n+1}) \\ &+ (-1)^n f(a_1, \cdots, a_{n-1}, a_n a_{n+1}) - (-1)^n f(a_1, \cdots, a_n) \cdot a_{n+1}.\end{aligned}$$

Finally, for $n = 0$, define $\delta p(a) = a \cdot p - p \cdot a$.

* Received May 16, 1951.

[1] This paper is based on a portion of the author's doctoral dissertation at Harvard University.

It is then easily proved (see [4], p. 60) that δ is a vector space homomorphism such that $\delta\delta = 0$, so that we have here a special case of the preceding general situation. In this case we denote Z^n by $Z^n(A, P)$, B^n by $B^n(A, P)$ etc.

This paper is concerned with several problems suggested by Hochschild. We consider first a problem proposed in [4]. After introducing for a given algebra the condition C_m: "All m-dimensional cohomology groups vanish," the statement is made (p. 58) "Theorem 3.1 implies that C_{m+1} is a consequence of C_m for $m \geq 1$. But it is an an open question whether or not C_m and C_{m+1} are equivalent."

We denote by K_m, $m = 0, 1, 2, \cdots$ the class of all algebras over F whose m-dimensional cohomology groups are all zero, or in other words cocycles are all coboundaries. (It is easily proved that K_0 is the null class.) One may then paraphrase the question concerning the condition C_m as follows: For $m \geq 0$ we have $K_m \subset K_{m+1}$; do we actually have $K_m = K_{m+1}$?

For $m = 0, 1, 2$ Hochschild has proved that the answer to this question is in the negative. The proof for $m = 0$ follows from his demonstration that the algebras in K_1 are a well-known non-null set, namely the set of algebras separable over F ([4], Theorem 4.1). For $m = 1$ an algebra (which we denote H') is produced such that $H' \varepsilon K_2 - K_1$ ([4], section 9). The case $m = 2$ is disposed of by first proving that adjoining an identity to an algebra does not affect its K-class ([5], section 2); then, letting H be the algebra formed by adjoining an identity to H', it is proved that the Kronecker product $H \times H \varepsilon K_3 - K_2$ ([6],Theorem 9.2).

Immediately following Hochschild's proof of this last result he states a conjecture which we shall call Conjecture 1.

CONJECTURE 1. *The n-fold Kronecker product* $H \times H \times \cdots \times H$ *does not belong to* K_n.

We prove Conjecture 1 in Prop. 6.1. There easily follows (Theorem 6.1) the answer to the first question raised, namely: $K_m \neq K_{m+1}$; for although the n-fold product $H \times H \times \cdots \times H$ does not belong to K_n, it does, as a consequence of Theorem 5.1, belong to K_{n+1}.

Conjecture 1 is followed in [6] by another conjecture:

CONJECTURE 2. *If* $A \varepsilon K_p - K_{p-1}$ *and* $B \varepsilon K_q - K_{q-1}$, *then the Kronecker product* $A \times B \varepsilon K_{p+q-1} - K_{p+q-2}$.

We shall consider in this paper the following two conjectures which together imply Conjecture 2.

CONJECTURE 2a. $A \varepsilon K_p, B \varepsilon K_q \Longrightarrow A \times B \varepsilon K_{p+q-1}$.

CONJECTURE 2b. $A \not\varepsilon K_p, \quad B \not\varepsilon K_q \Longrightarrow A \times B \not\varepsilon K_{p+q}$.

A special case of Conjecture 2a, namely that in which A and B have identities and $q = 2$, is a consequence of the following theorem in [6]:

THEOREM 9.1. *Let A and B be algebras with identity elements, and suppose that all two-dimensional cohomology groups of B are zero. Then for every $A \times B$ module Q and $n \geqq 1$, we have*

$$H^n[A, Z^0(B, C_1(A \times B, Q))] \approx H^{n+1}(A \times B, Q).$$

In [6] the proof of Theorem 9.1 is followed by another conjecture:

CONJECTURE 3. *If A and B have identities, and $B \varepsilon K_{p+1}$, then*

$$H^n[A, Z^0(B, C^p(A \times B, Q))] \approx H^{n+p}(A \times B, Q).$$

We prove Conjecture 3 in Theorem 4.1. There immediately follows (Theorem 5.1) that Conjecture 2a is true for algebras with identities. (Conjectures 2b and 2 are definitely not true for algebras A, B without identities; a counterexample will be exhibited in a subsequent paper on the classification of algebras by means of cohomology theory). For the case $p = 0$, A with an identity we prove Conjecture 2b in Cor. 5.1, while in the case $p = 1$, A and B with identities, a proof is given in Theorem 5.2. It is likely that Conjecture 2b is true for algebras A, B with identities, but a proof covering all cases remains to be found.

2. Notations and conventions. Except where otherwise indicated, the following notation and conventions will be assumed.

F: A field over which all algebras in the sequel are defined; all algebras in the sequel have identities.

$A \times B$: The Kronecker product of algebras A, B. If $1_\alpha, 1_\beta$ are the identities of A, B respectively, we identify $a \times 1_\beta$ with $a \varepsilon A$ and $1_\alpha \times b$ with $b \varepsilon B$ in situations where these identifications lead to no ambiguity; also, we denote $1_\alpha \times 1_\beta$ by 1.

S_i: The element $a_i \times b_i \varepsilon A \times B$.

${}_ra_s$: The sequence $a_r, a_{r+1}, \cdots, a_s$, $r \leqq s$.

L_p: The class of algebras $K_{p+1} - K_p$, $p = 0, 1, 2, \cdots$; note that $L_0 = K_1$.

3. The D-modules M_n and M'_n. $(n = 0, 1, 2, \cdots)$. Let D be an algebra, M a D-module. We shall find it useful to make the vector space $C^n(D, M)$ into a D-module in two ways.

(I) *For* $n > 0$, $C^n(D, M)$ *is made into a D-module* M_n *as follows: Let* $d, {}_1d_n \in D$, $g \in C^n(D, M)$. *Then we define* $d \cdot g$ *and* $g \cdot d$ *such that*

$$\{d \cdot g\}({}_1d_n) = d \cdot g({}_1d_n) \text{ and } \{g \cdot d\}({}_1d_n) = d \cdot g({}_1d_n) - \delta g(d, {}_1d_n).$$

(II) *For* $n > 0$, $C^n(D, M)$ *is made into a D-module* M'_n *as follows: Let* $d, {}_1d_n \in D$, $g \in C^n(D, M)$. *Then we define* $d \cdot g$ *and* $g \cdot d$ *such that*

$$\{d \cdot g\}({}_1d_n) = (-1)^n \delta g({}_1d_n, d) + g({}_1d_n) \cdot d \text{ and } \{g \cdot d\}({}_1d_n) = g({}_1d_n) \cdot d.$$

The definitions are completed by setting $M_0 = M'_0 = M$; to verify that M_n, M'_n actually are D-modules involves only straightforward computation.

The usefulness of these modules lies in the following two propositions.

PROPOSITION 3.1. *Given* $f \in C^{m+n}(D, M)$, $m \geq 0$, $n > 0$, *define* $\bar{f} \in C^m(D, M_n)$ *such that* $\bar{f}({}_1d_n) = f({}_1d_n)$ *for* $m = 0$, $\{\bar{f}({}_1d_m)\}({}_{m+1}d_{m+n}) = f({}_1d_{m+n})$ *for* $m > 0$. *Then* $\{\delta \bar{f}({}_1d_{m+1})\}({}_{m+2}d_{m+n+1}) = \delta f({}_1d_{m+n+1})$.

Proof. Note first that for $m > 0$,

$$\begin{aligned}
&\{\bar{f}({}_1d_m) \cdot d_{m+1}\}({}_{m+2}d_{m+n+1}) \\
&\qquad = d_{m+1} \cdot \{\bar{f}({}_1d_m)\}({}_{m+2}d_{m+n+1}) - \{\delta[\bar{f}({}_1d_m)]\}({}_{m+1}d_{m+n+1}) \\
&\qquad = d_{m+1} \cdot \{\bar{f}({}_1d_m)\}({}_{m+2}d_{m+n+1}) - d_{m+1} \cdot \{\bar{f}({}_1d_m)\}({}_{m+2}d_{m+n+1}) \\
&\qquad\qquad + f({}_1d_m, d_{m+1}d_{m+2}, {}_{m+3}d_{m+n+1}) + \cdots + (-1)^n f({}_1d_{m+n}) \cdot d_{m+n+1}.
\end{aligned}$$

Therefore, for $m > 0$,

$$\begin{aligned}
&\{\delta \bar{f}({}_1d_{m+1})\}({}_{m+2}d_{m+n+1}) \\
&= d_1 \cdot f({}_2d_{m+n+1}) - f(d_1d_2, {}_3d_{m+n+1}) + \cdots + (-1)^m f({}_1d_{m-1}, d_md_{m+1}, {}_{m+2}d_{m+n+1}) \\
&\qquad - (-1)^m \{\bar{f}({}_1d_m) \cdot d_{m+1}\}({}_{m+2}d_{m+n+1}) = \delta f({}_1d_{m+n+1}), \qquad \text{q. e. d.}
\end{aligned}$$

For $m = 0$ the result follows immediately from (I).

PROPOSITION 3.2. *Given* $f \in C^{m+n}(D, M)$, $m \geq 0$, $n > 0$, *define* $\bar{f} \in C^m(D, M'_n)$ *such that* $\bar{f}({}_1d_n) = (-1)^n f({}_1d_n)$ *for* $m = 0$, $\{\bar{f}({}_{n+1}d_{n+m})\}({}_1d_n) = (-1)^n f({}_1d_{m+n})$ *for* $m > 0$.

Then $\{\delta \bar{f}({}_{n+1}d_{m+n+1})\}({}_1d_n) = \delta f({}_1d_{m+n+1})$.

Proof. Note first that for $m > 0$,

$$\begin{aligned}
&\{d_{n+1} \cdot \bar{f}({}_{n+2}d_{n+m+1})\}({}_1d_n) \\
&\qquad = (-1)^n \{\delta[\bar{f}({}_{n+2}d_{n+m+1})]\}({}_1d_{n+1}) + \{\bar{f}({}_{n+2}d_{n+m+1})\}({}_1d_n) \cdot d_{n+1} \\
&\qquad = (-1)^n [d_1 \cdot [\{\bar{f}({}_{n+2}d_{n+m+1})\}({}_2d_{n+1})] - \cdots \\
&\qquad\qquad + (-1)^n \{\bar{f}({}_{n+2}d_{n+m+1})\}({}_1d_{n-1}, d_nd_{n+1})] \\
&\qquad = d_1 \cdot f({}_2d_{m+n+1}) - \cdots + (-1)^n f({}_1d_{n-1}, d_nd_{n+1}, {}_{n+2}d_{m+n+1}).
\end{aligned}$$

Therefore, for $m > 0$,

$$\{\delta\bar{f}({}_{n+1}d_{m+n+1})\}({}_1d_n)$$

$$= \{d_{n+1} \cdot \bar{f}({}_{n+2}d_{n+m+1})\}({}_1d_n) - (-1)^n f({}_1d_n, d_{n+1}d_{n+2}, d_{n+3}d_{n+m+1}) + \cdots$$
$$+ (-1)^n(-1)^{m+1} f({}_1d_{n+m}) \cdot d_{n+m+1} = \delta f({}_1d_{m+n+1}), \qquad \text{q. e. d.}$$

For $m = 0$ the result follows immediately from (II).

COROLLARY 3.1. $H^{m+n}(D, M) = H^m(D, M_n)$ *for* $m > 0$, $n \geq 0$.

COROLLARY 3.2. $H^{m+n}(D, M) = H^m(D, M'_n)$ *for* $m > 0$, $n \geq 0$.

4. The cohomology group of a Kronecker product.

LEMMA 4.1. *Suppose* $p > 0$, $f \in C^{p+1}(A \times B, M)$ *and that*

(i) *f is a coboundary on B, i. e. there exists* $g \in C^p(B, M)$ *such that* $\{f - \delta g\}({}_1b_{p+1}) = 0$.

(ii) $$i > p,\ S_i \in B \Longrightarrow \delta f({}_1S_{p+2}) = 0.$$

Then there exists $h \in C^p(A \times B, M)$ *such that*

$$i > p,\ S_i \in B \Longrightarrow \{f - \delta h\}({}_1S_{p+1}) = 0.$$

Proof. Adopting the notation

$[\delta f({}_1S_{p+2})]_k$: the sum of the first k terms of the expansion of $\delta f({}_1S_{p+2})$,

${}_l[\delta f({}_1S_{p+2})]$: the sum of the l-th term and its successors in the expansion of $\delta f({}_1S_{p+2})$,

a_{np}: the product $a_n a_{n+1} \cdots a_p$, $(n \leq p)$,

and defining the cochains

$$u({}_1S_p) = a_{1p} \cdot g({}_1b_p) - f(a_{1p}, {}_1b_p),$$
$$v({}_1S_p) = (-1)^p f({}_1S_p, 1),$$
$$f_r({}_1S_p) = f({}_1S_{r-1}, a_{rp}, {}_rb_p), \qquad (1 < r \leq p),$$

we derive the following relations:

(1). $$\delta u(S_1, b_2) = S_1 \cdot u(b_2) - u(S_1b_2) + u(S_1) \cdot b_2$$
$$= S_1 \cdot g(b_2) - S_1 \cdot f(1, b_2) - a_1 \cdot g(b_1b_2) + f(a_1, b_1b_2) + a_1 \cdot g(b_1) \cdot b_2 - f(a_1, b_1) \cdot b_2$$
$$= a_1 \cdot \delta g(b_1, b_2) - [S_1 \cdot f(1, b_2)] + f(a_1, b_1) \cdot b_2$$
$$= \delta f(a_1, b_1, b_2) + f(S_1, b_2) - [\delta f(S_1, 1, b_2) - f(S_1, 1) \cdot b_2]$$
$$= f(S_1, b_2) - f(S_1, 1) \cdot b_2. \qquad (p = 1)$$

$$
\begin{aligned}
(2)\quad \delta u({}_1S_p, b_{p+1}) &= S_1 \cdot u({}_2S_p, b_{p+1}) - u(S_1S_2, {}_3S_p, b_{p+1}) + \cdots \\
&\quad + (-1)^p u({}_1S_{p-1}, S_p b_{p+1}) - (-1)^p u({}_1S_p) \cdot b_{p+1} \\
&= S_1 a_{2p} \cdot g({}_2b_{p+1}) - a_{1p} \cdot g(b_1 b_2, {}_3b_{p+1}) + \cdots \\
&\quad + (-1)^p a_{1p} \cdot g({}_1b_{p-1}, b_p b_{p+1}) \\
&\quad - (-1)^p a_{1p} \cdot g({}_1b_p) \cdot b_{p+1} - S_1 \cdot f(a_{2p}, {}_2b_{p+1}) \\
&\quad + f(a_{1p}, b_1 b_2, {}_3b_{p+1}) - \cdots - (-1)^p f(a_{1p}, {}_1b_{p-1}, b_p b_{p+1}) \\
&\quad + (-1)^p f(a_{1p}, {}_1b_p) \cdot b_{p+1} \\
&= a_{1p} \cdot \delta g({}_1b_{p+1}) - S_1 \cdot f(a_{2p}, {}_2b_{p+1}) + \delta f(a_{1p}, {}_1b_{p+1}) \\
&\quad - a_{1p} \cdot f({}_1b_{p+1}) + f(a_{1p} b_1, {}_2b_{p+1}) \\
&= f(S_1 a_{2p}, {}_2b_{p+1}) - S_1 \cdot f(a_{2p}, {}_2b_{p+1}). \qquad (p > 1)
\end{aligned}
$$

$$
\begin{aligned}
(3).\quad \delta v({}_1S_p, b_{p+1}) &= S_1 \cdot v({}_2S_p, b_{p+1}) - v(S_1S_2, {}_3S_p, b_{p+1}) + \cdots \\
&\quad + (-1)^p v({}_1S_{p-1}, S_p b_{p+1}) - (-1)^p v({}_1S_p) \cdot b_{p+1} \\
&= (-1)^p [S_1 \cdot f({}_2S_p, b_{p+1}, 1) - f(S_1S_2, {}_pS_3, b_{p+1}, 1) + \cdots \\
&\quad + (-1)^p f({}_1S_{p-1}, S_p b_{p+1}, 1) - (-1)^p f({}_1S_p, 1) \cdot b_{p+1}] \\
&= (-1)^p [\delta f({}_1S_p, b_{p+1}, 1) + (-1)^p f({}_1S_p, b_{p+1}) - (-1)^p f({}_1S_p, b_{p+1}) \cdot 1 \\
&\quad - (-1)^p f({}_1S_p, 1) \cdot b_{p+1}] \\
&= f({}_1S_p, b_{p+1}) - f({}_1S_p, b_{p+1}) \cdot 1 - f({}_1S_p, 1) \cdot b_{p+1}.
\end{aligned}
$$

$$
\begin{aligned}
(4).\quad \delta f_p({}_1S_p, b_{p+1}) &= S_1 \cdot f_p({}_2S_p, b_{p+1}) - f_p(S_1S_2, {}_3S_p, b_{p+1}) + \cdots \\
&\quad - (-1)^p f_p({}_1S_{p-2}, S_{p-1}S_p, b_{p+1}) + (-1)^p f_p({}_1S_{p-1}, S_p b_{p+1}) \\
&\quad - (-1)^p f_p({}_1S_p) \cdot b_{p+1} \\
&= S_1 \cdot f({}_2S_p, 1, b_{p+1}) - f(S_1S_2, {}_3S_p, 1, b_{p+1}) + \cdots \\
&\quad - (-1)^p f({}_1S_{p-2}, S_{p-1}S_p, 1, b_{p-1}) \\
&\quad + (-1)^p f({}_1S_{p-1}, a_p, b_p b_{p+1}) - (-1)^p f({}_1S_{p-1}, a_p, b_p) \cdot b_{p+1} \\
&= \delta f({}_1S_p, 1, b_{p+1}) - (-1)^p f({}_1S_p, 1) \cdot b_{p+1} \\
&\quad + (-1)^p f({}_1S_{p-1}, a_p, b_p b_{p+1}) - (-1)^p f({}_1S_{p-1}, a_p, b_p) \cdot b_{p+1} \\
&= (-1)^p [f({}_1S_{p-1}, a_p, b_p b_{p+1}) - f({}_1S_p, 1) \cdot b_{p+1} - f({}_1S_{p-1}, a_p, b_p) \cdot b_{p+1}].
\end{aligned}
$$

$$
\begin{aligned}
(5).\quad \delta f_r({}_1S_p, b_{p+1}) &= S_1 \cdot f_r(S_2, b_{p+1}) - f_r(S_1S_2, {}_3S_p, b_{p+1}) + \cdots \\
&\quad - (-1)^r f_r({}_1S_{r-2}, S_{r-1}S_r, {}_{r+1}S_p, b_{p+1}) \\
&\quad + (-1)^r f_r({}_1S_{r-1}, S_r S_{r+1}, {}_{p+1}S_{r+2}) + \cdots - (-1)^p f_r({}_1S_p) \cdot b_{p+1} \\
&= S_1 \cdot f\ ({}_2S_r, a_{(r+1)p}, {}_{r+1}b_{p+1}) - f(S_1S_2, {}_3S_r, a_{(r+1)p}, {}_{r+1}b_{p+1}) + \cdots \\
&\quad - (-1)^r f({}_1S_{r-2}, S_{r-1}S_r, a_{(r+1)p}, {}_pb_{r+1}) \\
&\quad - (-1)^r f({}_1S_{r-1}, a_{rp}, b_r b_{r+1}, {}_{r+2}b_{p+1}) \\
&\quad + \cdots - (-1)^p f({}_1S_{r-1}, a_{rp}, {}_rb_p) \cdot b_{p+1} \\
&= [\delta f({}_1S_r, a_{(r+1)p}, {}_{r+1}b_{p+1})]_r - {}_{r+2}[\delta f({}_1S_{r-1}, a_{rp}, {}_rb_{p+1})].
\end{aligned}
$$

(6). Case 1. $p=2$. $\sum_{r=2}^{p}(-1)^r \delta f_r({}_1S_p, b_{p+1}) = \delta f_2({}_1S_2, b_3)$

$$= f(S_1, a_2, b_2 b_3) - f({}_1S_2, 1)\cdot b_3 - f(S_1, a_2, b_2)\cdot b_3$$
$$= -\delta f(S_1, a_2, {}_2b_3) + S_1 \cdot f(a_2, {}_2b_3) - f(S_1 a_2, {}_2b_3)$$
$$+ f({}_1S_2, b_3) - f({}_1S_2, 1)\cdot b_3$$
$$= S_1 \cdot f(a_{2p}, {}_2b_{p+1}) - f(S_1 a_{2p}, {}_2b_{p+1}) + f({}_1S_p, b_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1}.$$

(6). Case 2. $p>2$. $\sum_{r=2}^{p}(-1)^r \delta f_r({}_1S_p, b_{p+1})$

$$= \{\delta f_2 - \delta f_3 + \cdots + (-1)^p \delta f_p\}({}_1S_p, b_{p+1})$$
$$= [\delta f({}_1S_2, a_{3p}, {}_3b_{p+1})]_2 - {}_4[\delta f(S_1, a_{2p}, {}_2b_{p+1})]$$
$$- [\delta f({}_1S_3, a_{4p}, {}_4b_{p+1})]_3 + {}_5[\delta f({}_1S_2, a_{3p}, {}_3b_{p+1})]$$
$$+ [\delta f({}_1S_4, a_{5p}, {}_5b_{p+1})]_4 - {}_6[\delta f({}_1S_3, a_{4p}, {}_4b_{p+1})]$$
$$\vdots \qquad\qquad \vdots$$
$$+(-1)^p[\delta f({}_1S_{p-2}, a_{(p-1)p}, {}_{p-1}b_{p+1})]_{p-2} - (-1)^p{}_p[\delta f({}_1S_{p-3}, a_{(p-2)p}, {}_{p-2}b_{p+1})]$$
$$-(-1)^p[\delta f({}_1S_{p-1}, a_p, {}_pb_{p+1})]_{p-1} + (-1)^p{}_{p+1}[\delta f({}_1S_{p-2}, a_{(p-1)p}, {}_{p-1}b_{p+1})]$$
$$+(-1)^p(-1)^p[f({}_1S_{p-1}, a_p, b_p b_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1} - f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}]$$
$$= -{}_4[\delta f(S_1, a_{2p}, {}_2b_{p+1})]$$
$$- f(S_1, S_2 a_{3p}, {}_3b_{p+1}) \qquad + f({}_1S_2, S_3 a_{4p}, {}_4b_{p+1})$$
$$- f({}_1S_2, S_3 a_{4p}, {}_4b_{p+1}) \qquad + f({}_1S_3, S_4 a_{5p}, {}_5b_{p+1})$$
$$\vdots \qquad\qquad \vdots$$
$$- f({}_1S_{p-3}, S_{p-2} a_{(p-1)p}, {}_{p-1}b_{p+1}) + f({}_1S_{p-2}, S_{p-1} a_p, {}_pb_{p+1})$$
$$- (-1)^p[f({}_1S_{p-1}, a_p, {}_pb_{p+1})]_{p-1}$$
$$+ f({}_1S_{p-1}, a_p, b_p b_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1} - f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}$$
$$= S_1\cdot f(a_{2p}, {}_2b_{p+1}) - f(S_1 a_{2p}, {}_2b_{p+1}) + f(S_1, S_2 a_{3p}, {}_3b_{p+1})$$
$$- f(S_1, S_2 a_{3p}, {}_3b_{p+1}) + f({}_1S_{p-2}, S_{p-1} a_p, {}_pb_{p+1})$$
$$- f({}_1S_{p-2}, S_{p-1} a_p, {}_pb_{p+1}) + f({}_1S_p, b_{p+1}) - f({}_1S_{p-1}, a_p, b_p b_{p+1})$$
$$+ f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}$$
$$+ f({}_1S_{p-1}, a_p, b_p b_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1} - f({}_1S_{p-1}, a_p, b_p)\cdot b_{p+1}$$
$$= S_1\cdot f(a_{2p}, {}_2b_{p+1}) - f(S_1 a_{2p}, {}_2b_{p+1}) + f({}_1S_p, b_{p+1}) - f({}_1S_p, 1)\cdot b_{p+1}.$$

It may now be verified by straightforward computation that the following definition of h will satisfy the requirements of the lemma:

For $p=1$, $h(S_1)=\{2u-v\}(S_1)-u(S_1)\cdot 1$.

$$\text{For } p>1,\ h({}_1S_p)=\{2u-v+2\sum_{r=2}^{p}(-1)^r f_r\}({}_1S_p)$$
$$-\{u+\sum_{r=2}^{p}(-1)^r f_r\}({}_1S_p)\cdot 1.$$

This completes the proof of the lemma.

LEMMA 4.2. *Suppose* $p\geqq 0$, $B\,\varepsilon\,K_{p+1}$, $f\,\varepsilon\,C^{p+1}(A\times B,M)$, *and that* $i>p$, $S_i\,\varepsilon\,B\Longrightarrow \delta f({}_1S_{p+2})=0$. *Then there exists* $h\,\varepsilon\,C^p(A\times B,M)$ *such that* $i>p$, $S_i\,\varepsilon\,B\Longrightarrow\{f-\delta h\}({}_1S_{p+1})=0$.

Proof. The result is trivial for $p=0$, and follows immediately from Lemma 4.1 for $p>0$.

LEMMA 4.3. *Suppose* $p\geqq 0$, $r\geqq 1$, $B\,\varepsilon\,K_{p+1}$, $f\,\varepsilon\,C^{r+p+1}(A\times B,M)$ and *that*

(i) $$i>p,\ S_i\,\varepsilon\,B\Longrightarrow\delta f({}_1S_{r+p+2})=0,$$

(ii) $$i>p+1,\ S_i\,\varepsilon\,B\Longrightarrow f({}_1S_{r+p+1})=0.$$

Then there exists $h\,\varepsilon\,C^{r+p}(A\times B,M)$ *such that*

$$i>p,\ S_i\,\varepsilon\,B\Longrightarrow\{f-\delta h\}({}_1S_{r+p+1})=0.$$

Proof. We apply section **3**, setting $D=A\times B$, $n=r$ and (in Prop. 3.1) $m=p+1$. We have, then, a cochain $\bar f\,\varepsilon\,C^{p+1}(A\times B,M_r)$ such that

(a) $$\{\delta\bar f({}_1S_{p+2})\}({}_{p+3}S_{r+p+2})=\delta f({}_1S_{r+p+2}).$$

Now let N_r be the submodule of M_r consisting of all t such that

(b) $$i>p+1,\ S_i\,\varepsilon\,B\Longrightarrow t({}_{p+2}S_{r+p+1})=0,$$

(c) $$i>p+2,\ S_i\,\varepsilon\,B\Longrightarrow\{t\cdot S_{p+2}\}({}_{p+3}S_{r+p+2})=0.$$

We assert that $\bar f\,\varepsilon\,C^{p+1}(A\times B,N_r)$; for (b), in this case, follows immediately from (ii). To verify that f satisfies (a), we note first the following equality occurring in the proof of Prop. 3.1:

$$\{\bar f({}_1S_{p+1})\cdot S_{p+2}\}({}_{p+3}S_{r+p+2})=(-1)^p[\delta f({}_1S_{r+p+2})-S_1\cdot f({}_2S_{r+p+2})$$
$$+f(S_1S_2,{}_3S_{r+p+2})-\cdots+(-1)^p f({}_1S_p,S_{p+1}S_{p+2},{}_{p+3}S_{r+p+2}).$$

Now if $i > p + 2$ and $S_i \varepsilon B$, we have the first term in the brackets zero by (i) and the remaining terms in the brackets zero by (ii), q. e. d.

Next we show that if we replace f and M in Lemma 4. 1 by $\bar{f}$ and N_r respectively, then $\bar{f}$ satisfies the hypothesis of Lemma 4. 1. First observe that (i) and (a) imply that $\bar{f}$ is a cocycle on B. But then, since $B \varepsilon K_{p+1}$, it follows that $\bar{f}$ is a coboundary on B. Next, note that the requirement $i > p$, $S_i \varepsilon B \Longrightarrow \delta\bar{f}({}_1S_{p+2}) = 0$ also follows from (i) and (a).

We may therefore conclude that for $p > 0$ there exists $\bar{h} \varepsilon C^p(A \times B, N_r)$ such that $i > p$, $S_i \varepsilon B \Longrightarrow \{\bar{f} - \delta\bar{h}\}({}_1S_{p+1}) = 0$; the same conclusion is trivial for $p = 0$.

Now we define $h \varepsilon C^{r+p}(A \times B, M)$ such that for $p > 0$: $h({}_1S_{r+p}) = \{\bar{h}({}_1S_p)\}({}_{p+1}S_{r+p})$; for $p = 0$: $h({}_1S_r) = \bar{h}({}_1S_r)$. To show that h satisfies the conditions of the lemma, we assert first that $i > p + 1$, $S_i \varepsilon B \Longrightarrow \delta h({}_1S_{r+p+1}) = 0$; for we have $\delta h({}_1S_{r+p+1}) = \{\delta h({}_1S_{p+1})\}({}_{p+2}S_{r+p+1})$, and our assertion then follows from the definition of N_r.

Referring to (ii) we see that this proves the lemma for $i > p + 1$, $S_i \varepsilon B$. There remains to prove only the case $i = p + 1$, $S_i \varepsilon B$. But

$$\begin{aligned}\delta h({}_1S_p, b_{p+1}, {}_{p+2}S_{r+p+1}) &= \{\delta\bar{h}({}_1S_p, b_{p+1})\}({}_{p+2}S_{r+p+1}) \\ &= \{\bar{f}({}_1S_p, b_{p+1})\}({}_{p+2}S_{r+p+1}) = f({}_1S_p, b_{p+1}, {}_{p+2}S_{r+p+1}), \qquad \text{q. e. d.}\end{aligned}$$

LEMMA 4. 4. *Suppose* $r \geqq 0$, $p \geqq 0$, $B \varepsilon K_{p+1}$, $f \varepsilon C^{r+p+1}(A \times B, M)$ *and*

$$i > p,\ S_i \varepsilon B \Longrightarrow \delta f({}_1S_{r+p+2}) = 0.$$

Then there exists $h \varepsilon C^{r+p}(A \times B, M)$ *such that*

$$i > p,\ S_i \varepsilon B \Longrightarrow \{f - \delta h\}({}_1S_{k+p+1}) = 0.$$

Proof. We use induction on r. For $r = 0$, the lemma coincides with Lemma 4. 2. Suppose, then, that the lemma is true for $r = k \geqq 0$, and that $r = k + 1$. Replacing D in Prop. 3. 2 by $A \times B$, and setting $n = 1$, $m = r + p = k + p + 1$, we have defined a cochain $\bar{f} \varepsilon C^{k+p+1}(A \times B, M'_1)$ such that $\{\delta\bar{f}({}_2S_{k+p+3}\}(S_1) = \delta f({}_1S_{k+p+3})$. But then $\bar{f}$ satisfies the hypothesis of the lemma for $r = k$, so that there exists by the inductive hypothesis a cochain $\bar{h}_1 \varepsilon C^{k+p}(A \times B, M'_1)$ such that

$$i > p,\ S_i \varepsilon B \Longrightarrow \{f - \delta\bar{h}_1\}({}_1S_{k+p+1}) = 0.$$

We now increase the dimension of $\bar{h}_1$ by defining $h_1 \varepsilon C^{k+p+1}(A \times B, M)$ such that $h_1({}_1S_{k+p+1}) = \{\bar{h}_1({}_2S_{k+p+1})\}(S_1)$ for $k + p > 0$, and $h_1(S_1) = \bar{h}_1(S_1)$ for $k + p = 0$. Then $\delta h_1({}_1S_{k+p+2}) = - \{\delta h_1({}_2S_{k+p+2})\}(S_1)$.

The cochain h_1 does not quite satisfy our requirements, since we now have only: $i > p+1, S_i \varepsilon B \Longrightarrow \{f - \delta h_1\}({}_1S_{k+p+2}) = 0$. However, this result means that $f - \delta h_1$ satisfies the hypotheses of Lemma 4. 3 for $r = k+1$, with f replaced by $f - \delta h_1$. There exists therefore a cochain $h_2 \varepsilon C^{k+p+1}(A \times B, M)$ such that $i > p$, $S_i \varepsilon B \Longrightarrow \{f - \delta h_1) - \delta h_2\}({}_1S_{k+p+2}) = 0$.

Clearly, $h = h_1 + h_2$ satisfies the requirements of the lemma.

LEMMA 4. 5. *Suppose* $r, p \geqq 0$, $B \varepsilon K_{p+1}$, $\bar{f} \varepsilon C^{r+1}(A \times B, M'_p)$ *and that*

$$i > p,\ S_i \varepsilon B \Longrightarrow \delta\bar{f}({}_{p+1}S_{p+r+2}) = 0.$$

Then there exists $\bar{h} \varepsilon C^r(A \times B, M'_p)$ *such that*

$$i > p,\ S_i \varepsilon B \Longrightarrow \{\bar{f} - \delta\bar{h}\}({}_{p+1}S_{r+p+1}) = 0.$$

Proof. If $p = 0$ the lemma is a special case of Lemma 4. 4. Suppose, then, that $p > 0$. We define f as in Prop. 3. 2, replacing D by $A \times B$ and setting $m = r+1$, $n = p$. Then we have $\{\delta\bar{f}({}_{p+1}S_{p+r+2})\}({}_1S_p) = \delta f({}_1S_{r+p+2})$. Therefore $i > p$, $S_i \varepsilon B \Longrightarrow \delta f({}_1S_{r+p+2}) = 0$.

But then f satisfies the hypothesis of Lemma 4. 4. Therefore there exists $h \varepsilon C^{r+p}(A \times B, M)$ such that $i > p$, $S_i \varepsilon B \Longrightarrow \{f - \delta h\}({}_1S_{r+p+1}) = 0$.

The cochain $\bar{h}$ defined as follows will now satisfy the requirements of the lemma: for $r > 0$, $\{\bar{h}({}_{p+1}S_{r+p})\}({}_1S_p) = (-1)^p h({}_1S_{r+p})$; for $r = 0$, $\bar{h}({}_1S_p) = (-1)^p h({}_1S_p)$.

THEOREM 4. 1. $B \varepsilon K_{n+1} \Longrightarrow H^{m+n}(A \times B, M) \cong H^m[A, Z^0(B, M'_n)]$, $(m > 0, n \geqq 0)$.

Proof. In Lemma 4. 5, change the notation by replacing $\bar{f}$ by f, $\bar{h}$ by g, M'_p by P and $r+1$ by m. The result will be Lemma 3 of [5], (p. 574). As an immediate consequence of this lemma Hochschild proves ([5], Theorem 6, p. 575) the following result: $H^m[A, Z^0(B, M'_p)] \cong H^m(A \times B, M'_p)$, $(m > 0, p \geqq 0)$. Theorem 4. 1 now follows from Cor. 3. 2.

5. The *K*- and *L*-classes of a Kronecker product.

THEOREM 5.1. $A \varepsilon K_m$, $B \varepsilon K_p \Longrightarrow A \times K_{m+p-1}$ $(m, p > 0)$.

Proof. Let M be any $A \times B$ module, $p = n+1$. Then the hypothesis of Theorem 4. 1 is satisfied, so that $H^{m+n}(A \times B, M) \cong H^m[A, Z^0(B, M'_n)]$. But since $A \varepsilon K_m$ we have that $H^m[A, Z^0(B, M'_n)] = 0$. Therefore $H^{m+n}(A \times B, M) = 0$, i. e. $A \times B \varepsilon K_{m+n} = K_{m+p-1}$, q. e. d.

With reference to determining a lower bound for the *K*-class of $A \times B$,

only the two simplest cases are settled here. The result for the simplest case is given in Cor. 5. 1. Lemma 5. 1 introduces a general construction which seems to yield a result (Theorem 5.2) only for the next case.

PROPOSITION 5. 1. *Let M be any B-module (B with or without an identity). We make P, the underlying vector space of $A \times M$, into an $A \times B$ module by the following definition:*

$$(a_1 \times b) \cdot (a_2 \times m) = a_1 a_2 \times b \cdot m, \quad (a_2 \times m) \cdot (a_1 \times b) = a_2 a_1 \times m \cdot b,$$

where $a_1, a_2 \varepsilon A$, $b \varepsilon B$ and $m \varepsilon M$ are elements of a pre-chosen basis for A, B, M respectively. This definition is then extended linearly to cover all of $A \times B$ and P.

Now let ϕ be the homomorphism of $C^k(B, M)$ into $C^k(A \times B, P)$ defined such that $\{\phi f\}({}_1S_k) = a_{1k} \times f({}_1b_k)$ for $k > 0$ and $\phi(m) = 1 \times m$ for $k = 0$.

Then ϕ induces an isomorphism of $H^k(B, M)$ into $H^k(A \times B, P)$, $(k \geq 0)$.

Proof. First we show that $\phi\delta = \delta\phi$. For $k > 0$,

$$\begin{aligned}\{\delta(\phi f)\}({}_1S_{k+1}) &= S_1 \cdot \phi f({}_2S_{k+1}) - \phi f(S_1S_2, {}_3S_{k+1}) + \cdots \\ &\quad + (-1)^k \phi f({}_1S_{k-1}, S_kS_{k+1}) - (-1)^k \phi f({}_1S_k) \cdot S_{k+1} \\ &= a_{1k} \times [b_1 \cdot f({}_2b_{k+1}) - f(b_1b_2, {}_3b_{k+1}) + \cdots \\ &\quad + (-1)^k f({}_1b_{k-1}, b_kb_{k+1}) - (-1)^k f({}_1b_k) \cdot b_{k+1}] \\ &= a_{1k} \times \delta f({}_1b_{k+1}) = \{\phi(\delta f)\}({}_1S_{k+1}),\end{aligned}$$

proving our assertion for $k > 0$.

For $k = 0$: $\delta\{(\phi m)\}(S) = S \cdot \phi m - \phi m \cdot S = (a \times b) \cdot (1 \times m) - (1 \times m) = a \times \delta m(b) = \{\phi(\delta m)\}(S)$, proving our assertion for $k = 0$.

Now consider the mapping $f \to f \bmod B^k(A \times B, P)$. Since $\phi\delta = \delta\phi$, this mapping carries cocycles into cocycles and coboundaries into coboundaries. Thus all that remains to be proved is that the kernel of this mapping consists *only* of $B^k(B, M)$, in other words that if $\phi f = \delta g$, then there exists $h \varepsilon C^k(B, M)$ such that $f = \delta h$.

Let $1 = u_0, u_1 \cdots, u_r$ be a basis for A. We define a projection π of $A \times M$ into $1 \times M$ as follows. If $c \varepsilon A \times M$ and $c = \sum_{i=0}^{r} u_i \times m_i$, let $\pi c = 1 \times m_0$. Now, supposing $\phi f = \delta g$, let $1 \times h = \pi g$. Then

$$\phi f({}_1b_k) = 1 \times f({}_1b_k) = \delta g({}_1b_k) = 1 \times \delta h({}_1b_k).$$

Therefore $f = \delta h$, q. e. d.

COROLLARY 5.1.[2] $B \notin K_n \Longrightarrow A \times B \notin K_n$, ($B$ with or without identity, $n \geq 0$).

COROLLARY 5.2. $A \in L_0$, $B \in L_n \Longrightarrow A \times B \in L_n$, $(n \geq 0)$.

Proof of Corollary 5.2. $A \times B \notin K_n$ follows from Cor. 5.1; by Theorem 5.1 we have $A \times B \in K_{1+n+1-1} = K_{n+1}$, q. e. d.

We now define an algebraic "cup product" analogous to the cup product defined for groups by Eilenberg and MacLane. ([2], section 4).

Definition 5.1. Let A, B be algebras with or without identities, $f \in C^r(A, M)$, $g \in C^s(B, N)$, $(r, s > 0)$.[3]. Consider the vector space underlying $M \times N$ as an $A \times B$ module such that $(a \times b) \cdot (m \times n) = (a \cdot m) \times (b \cdot n)$, $(m \times n) \cdot (a \times b) = (m \cdot a) \times (n \cdot b)$.

Then the cup product $f \cup g \in C^{r+s}(A \times B, M \times N)$ is defined as follows:

$$\{f \cup g\}({}_1s_{r+s}) = [f({}_1a_r) \cdot a_{(r+1)(r+s)}] \times [b_{1r} \cdot g({}_{r+1}b_{r+s})].$$

It may be verified by straightforward computation that the cup product is associative, and we have

PROPOSITION 5.2. $\delta(f \cup g) = \delta f \cup g + (-1)^r f \cup \delta g$.

As a consequence of the associativity of the cup product we may extend Definition 5.1 inductively to any finite sequence of cochains; we note also that Proposition 5.2 implies that the cup product of cocycles is a cocycle and that the cup product of a cocycle and a coboundary (in either order) is a coboundary.

LEMMA 5.1. *Let $f \in Z^r(A, M)$. $g \in Z^s(B, N)$, $A \times B \in K_{r+s}$ $(r, s > 0)$, and suppose the identity elements of A, B act as identity operators on M, N respectively, and that f is "normalized" i. e. $1 \leq i \leq r$, $a_i = 1 \Longrightarrow f({}_1a_r) = 0$.*

Then given sequences $\alpha = {}_1a_r \in A$, $\beta = {}_1b_s \in B$, there exist cochains $f_\alpha \in C^{s-1}(B, M \times N)$ and $f_\beta \in C^{r-1}(A, M \times N)$ such that

$$f({}_1a_r) \times g({}_1b_s) = \delta f_\beta({}_1a_r) + \delta f_\alpha({}_1b_s).$$

Proof. Let $h = f \cup g$. By the remark preceding this lemma, h is a cocycle; hence, since $A \times B \in K_{r+s}$, h is a coboundary. Therefore there exists $\bar{h} \in C^{r+s-1}(A \times B, M \times N)$ such that $h = \delta\bar{h}$. Now consider

---

[2] For algebras A, B both with identities, this result is derived in Hochschild [5] as a consequence of Theorem 7, *ibid.*

[3] Although we shall not find it necessary to do so in the sequel, Def. 5.1 may be extended to include the cases $r = 0$, $s = 0$ also. In fact the cochain ϕf of Prop. 5.1 may be considered as a cup product in which $r = 0$, $s > 0$.

(I)

$$
\begin{array}{llll}
h({}_1a_r,{}_1b_s) & = a_1\cdot\bar{h}({}_2a_r,{}_1b_s) & -\cdots+(-1)^r\bar{h}({}_1a_{r-1},a_r\times b_1,{}_2b_s)-\cdots \\
-h({}_1a_{r-1},b_1,a_r,{}_2b_s) & = -a_1\cdot\bar{h}({}_2a_{r-1},b_1,a_r,{}_2b_s)+ & \cdots-(-1)^r\bar{h}({}_1a_{r-1},a_r\times b_1,{}_2b_s)-\cdots \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot
\end{array}
$$

The arguments of h in the first column of (I) are to consist of the $(r+s)!/r!s!$ permutations of the sequence $({}_1a_r,{}_1b_s)$ which do not alter the order of the a's or the order of the b's. The signs at the extreme left of the rows of (I) are $+$ or $-$ according as the argument of h is an even or an odd permutation of $({}_1a_r,{}_1b_s)$.

Now we define $f_\beta=\bar{h}({}_1b_s)$ for $r=1$ and $f_\alpha=(-1)^r\bar{h}({}_1a_r)$ for $s=1$. For $r,s>1$ we let $f_\beta({}_1a_{r-1})=\bar{h}({}_1a_{r-1},{}_1b_s)-\cdots$, where the terms on the right are of the form $\pm\bar{h}({}_1S_{r+s-1})$, with the sequences $({}_1S_{r+s-1})$ consisting of all permutations of $({}_1a_{r-1},{}_1b_s)$ which do not alter the order of the a's or the order of the b's; the sign preceding $\bar{h}$ is to be $+$ or $-$ according as the argument of h is an even or an odd permutation of $({}_1a_{r-1},{}_1b_s)$. Finally, let $f_\alpha=(-1)^r[\bar{h}({}_1a_r,{}_1b_{s-1})-\cdots]$, with a similar convention as to the arguments of $\bar{h}$.

We assert that the sum of the right hand sides of the equation in (I) is $\delta f_\beta({}_1a_r)+\delta f_\alpha({}_1b_s)$. For firstly, if a summand on the right has an argument in which a term of the type $a_i\times b_j$ appear, then that summand appears precisely twice on the right and with opposite signs, so that all such summands disappear in adding the rows of (I); secondly, the remaining terms each occur precisely once and are clearly the terms of the expansions of $\delta f_\alpha({}_1b_s)$ and $\delta f_\beta({}_1a_s)$.

Finally, from the hypothesis of the lemma and the definition of h, it follows that $h({}_1a_r,{}_1b_s)=f({}_1a_r)\times g({}_1b_s)$, while all the other terms on the left hand sides of the equations in (I) are zero. The sum of the left hand sides of the equations in (I) will then be $f({}_1a_r)\times g({}_1b_s)$, which completes the proof of the lemma.

LEMMA 5.2. *If $A\not\varepsilon K_1$, then there exist an A-module M, $f\,\varepsilon\,Z^1(A,M)$ and $a_0\,\varepsilon\,A$ such that the identity of A acts as an identity operator on M, $f(a_0)\neq 0$ and $m\,\varepsilon\,M\Longrightarrow a_0\cdot m=m\cdot a_0$.*

Proof. Case 1. $R\neq 0$ is the radical of A and A/R is separable. Then $A=A/R+R$ as a supplementary sum, i.e. each element $a\,\varepsilon\,A$ may be uniquely expressed as $a=a'+r_a$, where $a'\,\varepsilon\,A/R$ and $r_a\,\varepsilon\,R$. Let π be the

projection of A onto R such that $\pi(a) = r_a$, let $M = R/R^2$, denote r/R^2 by $\bar{r}$, where $r \varepsilon R$, and define $a \cdot \bar{r} = \overline{ar}$, $\bar{r} \cdot a = \overline{ra}$, $f(a) = \overline{\pi a}$.

Then $1 \cdot \bar{r} = \bar{r} \cdot 1 = \bar{r}$; if we let a_0 be any element of $R - R^2$ then $f(a_0) = \overline{\pi a_0} = \bar{a}_0 \neq \bar{0}$; $\delta f(a'_1 + r_1, a'_2 + r_2) = \overline{a'_1 r_2 - (a'_1 r_2 + r_1 a'_2) + r_1 a'_2} = \bar{0}$, so that $f \varepsilon Z^1(A, M)$; finally, $a_0 \cdot \bar{r} = \bar{0} = \bar{r} \cdot a_0$. Thus all the requirements of the lemma are satisfied.

Case 2. $R = 0$.

In this case A is semisimple and inseparable. Let A_1 be a simple component of A of dimension m over its center C. It is shown in Hochschild [4], Lemma 4. 1 that we may make the matrix algebra C_m into an A-module such that the identity of A acts as an identity operator on M, and such that there exists $a_0 \varepsilon C$ and $f \varepsilon Z^1(A, M)$ with $f(a_0) \neq 0$ and $a_0 \cdot m = m \cdot a_0$ for all $m \varepsilon M$. By allowing the other simple components of A to annihilate M and defining f to be zero on the other simple components of A, we make M into an A-module and extend f to A. The requirements of the lemma are then satisfied.

Case 3. $R \neq 0$ and A/R is inseparable.

Since A/R is semisimple, there exists by Case 2 an A/R module $\bar{M}$ and a cocycle $\bar{f} \varepsilon Z^1(A/R, \bar{M})$ satisfying the requirements of the lemma, with A/R, $\bar{M}, \bar{f}$ replacing A, M, F respectively. Letting $\bar{a}$ represent a/R, we may achieve the desired result by defining the vector space underlying M to coincide with that underlying $\bar{M}$, and for $m \varepsilon M$, $a \varepsilon A$ defining $a \cdot m = \bar{a} \cdot m$, $m \cdot a = m \cdot \bar{a}$, $f(a) = \bar{f}(\bar{a})$.

Theorem 5. 2. $A \not\varepsilon K_1$, $B \not\varepsilon K_n \Longrightarrow A \times B \not\varepsilon K_{n+1}$, $(n > 0)$.

Proof. Let M, f, a_0 be as in Lemma 5. 2, $g \varepsilon Z^n(B, N) - B^n(B, N)$, ${}_1b_n \varepsilon B$ and suppose $A \times B \varepsilon K_{n+1}$. Then by Lemma 5. 1 there exist cochains $g_\alpha \varepsilon C^n(B, M \times N)$, $f_\beta \varepsilon C^0(A, M \times N)$ such that $f(a) \times g({}_1b_n) = \delta f_\beta(a) + \delta g_\alpha({}_1b_n)$.

We may assume (cf. Hochschild [5], section 1) that the identity element of B acts as an identity operator on N. Consequently Lemma 5. 1 implies that $\delta f_\beta(a_0) = 0$. Now if we let $f(a_0) \times h({}_1b_{n-1})$ be a projection of $g_\alpha({}_1b_{n-1})$ into $f(a_0) \times N$, we have $g = \delta h$, a contradiction which proves the theorem.

Corollary. $A \varepsilon L_1$, $B \varepsilon L_n \Longrightarrow A \times B \varepsilon L_{n+1}$, $(n \geq 0)$.

6. The existence of algebras in L_n, $(n \geq 0)$.

We now prove that none of the classes L_n is null. L_0 is not null, of

course, since $L_0 = K_1$ is the class of algebras separable over F. As for L_1, it is proved by Hochschild ([5], section 9) that $H' \varepsilon K_2$ where H' is an algebra with basis a, r such that a is a left identity and r a left annihilator. Since every separable algebra has an identity, it follows that $H' \not\varepsilon K_1$. Let H be the algebra formed by adjoining an identity 1 to H'. Then $H \varepsilon L_1$, since adjoining an identity to an algebra does not affect its K-class (cf. Hochschild [5], section 2).

PROPOSITION 6.1.

$$H_1 = H_2 = \cdots = H_n = H \Longrightarrow H_1 \times H_2 \times \cdots \times H_n \not\varepsilon K_n.$$

Proof. Let R (with basis r) denote the radical of H. Let $R_1 = R_2 = \cdots = R_n = R$. We make $R_1 \times \cdots \times R_n$ into an $H_1 \times \cdots \times H_n$ module by defining $(h_1 \times \cdots \times h_n) \cdot (r_1 \times \cdots \times r_n) = h_1 r_1 \times \cdots \times h_n r_n$, $(r_1 \times \cdots \times r_n) \cdot (h_1 \times \cdots \times h_n) = r_1 h_1 \times \cdots \times r_n h_n$. Next define $f \varepsilon C^1(H, R)$ such that $f(1) = f(a) = 0$, $f(r) = r$. Then it is easily verified (either directly or from Lemma 5.2, Case 1) that $f \varepsilon Z^1(H, R)$.

Now let g be the cup product $f_1 \cup \cdots \cup f_n$, where $f_1 = \cdots = f_n = f$. Then by the remark following Prop. 5.2, g is a cocycle. If for $h_{ij} \varepsilon H$ we denote $h_{i1} \times \cdots \times h_{in}$ by U_i and $[h_{1i} \cdots h_{(i-1)i}][f(h_{ii})][h_{(i+1)i} \cdots h_{ni}]$ by V_i, then $g({}_1U_n) = V_1 \times \cdots \times V_n$.

In particular, if $h_{ii} = r$ and $h_{ij} = 1$ for $i \neq j$, then we denote U_i by W_i. Now, supposing that $g = \delta h$, consider the following set of equations (for convenience we let $W = W_1 W_2 = W_2 W_1$):

(II)

$$g(W_1, W_2, {}_3W_n) = W_1 \cdot h(W_2, {}_3W_n) - h(W, {}_3W_n) + \cdots + (-1)^n h(W_1, W_2, {}_3W_{n-1}) \cdot W_n$$
$$-g(W_2, W_1, {}_3W_n) = - W_2 \cdot h(W_1, {}_3W_n) + h(W, {}_3W_n) - \cdots - (-1)^n h(W_2, W_1, {}_3W_{n-1}) \cdot W_n$$
$$\begin{matrix} \cdot & \quad & \cdot & \quad & \cdot & \quad & \cdot \\ \cdot & & \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot & & \cdot \end{matrix}$$

The arguments of g in the first column of (II) are to consist of the $n!$ permutations of the terms of the sequence $({}_1W_n)$. The signs at the extreme left of the rows of (II) are to be $+$ or $-$ according as the argument of g is an even or an odd permutation of $({}_1W_n)$.

We now note that each term in the first and last columns on the right hand side of the equations in (II) is zero, since each of these terms is equal to a Kronecker product with a factor in $R^2 = 0$. Secondly, note that each term in the columns between those just mentioned occurs precisely twice in

(II), but with opposite signs. The sum of the right hand terms of (II) is therefore zero. On the left hand side of the equations of (II), it is easily verified that all terms are zero except $g({}_1W_n) = r \times r \times \cdots \times r$.

We are therefore led by the assumption $g = \delta h$ to the contradiction $r \times r \times \cdots \times r = 0$. Thus g is a cocycle but not a coboundary; hence $H_1 \times \cdots \times H_n \not\in K_n$, q. e. d.

THEOREM 6. 1. *L_n is not null, $(n \geq 0)$.*

Proof. We have already noted that $L_0 = K_1$ is not null. For $n > 0$, as an immediate consequence of Theorem 5. 1 and Prop. 6. 1, it follows that the algebra $H_1 \times \cdots \times H_n \varepsilon L_n$, where $H_1 = \cdots = H_n = H$.

UNIVERSITY OF MASSACHUSETTS.

BIBLIOGRAPHY

[1] S. Eilenberg, "Topological methods in abstract algebra," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 3-37.

[2] S. Eilenberg and S. MacLane, 'Cohomology theory in abstract groups. I," *Annals of Mathematics*, vol. 48 (1947), pp. 51-78.

[3] ———, "Cohomology theory in abstract groups. II," *ibid.*, vol. 48 (1947), pp. 326-341.

[4] G. Hochschild, "On the cohomology groups of an associative algebra," *ibid.*, vol. 46 (1945), pp. 58-67.

[5] ———, "On the cohomology theory for associative algebras," *ibid.*, vol. 47 (1946), pp. 568-579.

[6] ———, "Cohomology and representations of associative algebras," *Duke Mathematical Journal*, vol. 14 (1947), pp. 921-948.

ON RELATED PERIODIC MAPS.*

By E. E. Floyd.

1. Introduction. Consider a class of periodic maps defined on a topological space X. We are concerned with special cases of the following problem. Suppose the maps of the class are all related in some specified fashion. Are there, then, any implied relationships between the fixed point sets of the maps of the class?

A notable example of a problem of this sort has been solved recently by S. D. Liao [5]. If X is a finite dimensional compact Hausdorff space which has the homology groups of an n-sphere over the group I_p of integers mod p with p prime, and if T is periodic of period p on X, then, as P. A. Smith has proved ([8], p. 366), the fixed point set L has the homology groups of a r-sphere for some $-1 \leq r \leq n$. Liao settled a problem proposed by Smith by proving that if X also has finitely generated integral cohomology groups, then $n - r$ is even or odd according as T is orientation preserving or orientation reversing.

In section 1, we generalize Liao's result by proving that if X is a finite dimensional compact Hausdorff space with finitely generated integral cohomology groups, and if T is periodic of prime power period p^a on X, then the Lefschetz fixed point number of T is equal to the Euler characteristic of L (defined using I_p as coefficient group). We also extend a result of Smith ([9], p. 162) concerning the non-existence of certain types of periodic maps of arbitrarily large period on n-manifolds with negative Euler characteristic. The methods of this section depend heavily on recent results of Liao [5] and of the author [4] which in turn are based on the special homology groups of Smith [8].

In section 2, we consider a periodic map T of prime power period q^a and then consider the class of all periodic maps T_1 of the same period which are "sufficiently close" to T. Under these circumstances, we prove that the fixed point set L_1 of T_1 is close to L in the sense of Begle's metric [1] induced by the regular convergence introduced by Whyburn [11].

The author has read a pre-publication copy of Mr. Liao's paper [5], and wishes to thank Mr. Liao for that privilege.

* Received August 24, 1951; revised October 25, 1951.

2. The Lefschetz fixed point number of T. A periodic map on a space X generates a periodic linear isomorphism on the rational homology groups of X. We require later in the section an analysis of the latter. We dispose of this first, using a procedure similar to one used by Smith ([9], pp. 161-162) for a similar purpose.

Suppose V is a finite dimensional vector space over the rationals R. If W is a subspace of V, let dW denote the dimension of W. Let T be a linear transformation on V with $T^p =$ identity. There are associated with T the linear transformations $\sigma = 1 + T + \cdots + T^{p-1}$ and $\tau = 1 - T$. Clearly $\sigma\tau = \tau\sigma = 0$. We use the following preliminary remark (cf. [5], 4.11).

(2.1) *Image* $\sigma =$ *kernel* τ.

If m is a matrix presentation of T, then we call its characteristic equation $f(t)$ the characteristic equation of T. The characteristic roots of T are p-th roots of unity, for if $| m - \lambda I | = 0$, then $0 = | m^p - \lambda^p I | = (1 - \lambda^p)^{dV}$. Moreover, if no T^i, $0 < i < p$, has non-zero fixed points, then every characteristic root λ is a primitive p-th root of unity. For if $\lambda^l = 1$, then $| I - m^l | = | \lambda^l I - m^l | = 0$. Hence there exists $x \in V$, $x \neq 0$, with $T^l x = x$. But then $l = p$, so λ is a primitive p-th root.

Since $f(t)$ has rational coefficients and all its roots are p-th roots of unity, then $f(t) = f_{s_1}(t) \cdots f_{s_k}(t)$ where $f_{s_i}(t)$ is the cyclotomic equation of degree $\phi(s_i)$, and ϕ is Euler's ϕ-function, whose roots are the primitive s_i-th roots of unity. Moreover it may be seen that s_i divides p. In the following, we use $V(S)$ to represent the fixed point set of the linear transformation S.

(2.2) *Let T be a linear transformation on the finite-dimensional rational vector space V with $T^p =$ identity. Then*

(a) *if p is prime, there exists a non-negative integer k with $dV = dV(T) + k(p-1)$; moreover, trace $T = dV(T) - k$;*

(b) *if $p = q^a$ where q is prime and $a > 1$, then trace $T =$ trace $T \mid V(T^{q^{a-1}})$.*

Proof. To prove (a), decompose V into $V(T) \oplus V_1$, where $T(V_1) = V_1$ (cf. the proof of (2.1)). The characteristic equation of $T \mid V_1$ has as roots only primitive p-th roots of unity. Hence its characteristic equation is of the form $(f_p(t))^k$. Since the degree of $f_p(t)$ is $p-1$, $dV = dV(T) + k(p-1)$. The trace of $T \mid V_1$ is then $k(\alpha_1 + \cdots + \alpha_{p-1})$, where the α_i's are the primitive p-th roots of unity. Hence the trace of $T \mid V_1$ is $-k$. So (a) follows.

To prove (b), decompose V into $V(T^{q^{a-1}}) \oplus V_1$, where $T(V_1) = V_1$. Then the characteristic equation of $T \mid V_1$ is of the form $(f_p(t))^k$, and the trace of $T \mid V_1$ is $k(\alpha_1 + \cdots + \alpha_{\phi(p)})$, where the α_i's are the primitive p-th roots of unity. It may then be seen that the trace of $T \mid V_1$ is 0. So (b) follows.

Suppose now that X is a compact Hausdorff space, and let T be a map of X into X. Let $H_n(X; F)$ denote the Čech homology group of X over the field F, and T_{*n} the induced linear transformation on $H_n(X; F)$. Define $\chi(X; F) = \Sigma(-1)^i dH_i(X; F)$, in case the right hand side is defined and finite, and call $\chi(X; F)$ the Euler characteristic of X over F. Also define $\alpha(T; F) = \Sigma(-1)^i$ trace T_{*i}, in case $\chi(X; F)$ exists, and call $\alpha(T; F)$ the Lefschetz fixed point number of T over F ([6], p. 319).

We suppose now that X is a finite dimensional compact Hausdorff space with finitely generated integral Čech cohomology groups. Let T denote a periodic map on X of prime period p. Let L denote the fixed point set of T, and Y the orbit decomposition space of T. We have occasion to use the following recent results. Of these, (2.3), (2.4), and (2.5) are due to Liao [5], and (2.6) to the author [4].

(2.3) (*Liao*). *Y has finitely generated cohomology groups.*

Liao ([5], Theorem 5.5) has given a proof for this in case X has the groups of an n-sphere over I_p. The proof used the extra assumption only to insure that L has finitely generated groups over I_p. Since this is true in the general case ([4], Theorem 4.2), the proof then holds.

(2.4) (*Liao*). $\chi(X; I_p) = \chi(X; R)$, $\chi(Y; I_p) = \chi(Y; R)$ ([5], Theorem 2.8).

(2.5) *If $\eta: X \to Y$ denotes the orbit decomposition map, then η_* maps $[x \mid x \in H_n(X; R), T_* x = x]$ isomorphically onto $H_n(Y; R)$.*

This result is more or less implicit in the work of Liao (cf. [5], 4.3, 4.11, 4.13). Because of its importance here, we outline, using the notation of [5; §4], a direct argument. For each $b_{s\lambda} \in C_s(0(K_\lambda, T_\lambda); R)$, let $a_{s\lambda} \in C_s(K_\lambda, R)$ be such that $\eta_\lambda(a_{s\lambda}) = b_{s\lambda}$. Define $\xi_\lambda(b_{s\lambda}) = \sigma_\lambda a_{s\lambda}$. It may be verified that ξ_λ is uniquely defined, that $\partial\xi_\lambda = \xi_\lambda\partial$, and that $\pi_{\mu\lambda}\xi_\mu = \xi_\lambda\pi_{0\mu\lambda}$. Moreover, $\xi_\lambda\eta_\lambda = \sigma_\lambda$, and $\eta_\lambda\xi_\lambda(b_{s\lambda}) = pb_{s\lambda}$. Hence there is induced $\xi: H_s(0(X, T); R) \to H_s(X; R)$ with $\eta\xi(x) = px, x \in H_s(0(X, T); R)$, $\xi\eta(x) = \sigma(x), x \in H_s(X; R)$. Since $\eta\xi$ is an isomorphism onto, η maps image ξ isomorphically onto $H_s(0(X, T); R)$. Since η is onto and $\xi\eta = \sigma$, we have image ξ = image σ. But by (2.1) image σ = kernel τ. The assertion follows.

(2.6) $\chi(X;I_p)+(p-1)\chi(L;I_p)=p\chi(Y;I_p)$. [4].

We are now in a position to prove the main theorem of this section.

(2.7) THEOREM. *Let X be a finite dimensional compact Hausdorff space with finitely generated integral Čech cohomology groups. Let T be a periodic map on X of period q^a, q prime. Let L be the fixed point set of T. Then $\alpha(T;R)=\chi(L;I_q)$.*

Proof. We prove the theorem first for $a=1$. Consider $T_{*n}\colon H_n(X;R)\to H_n(X;R)$. According to (2.5), the fixed point set of T_{*n} is isomorphic to $H_n(Y;R)$. Hence by (2.5),

$$dH_n(X;R)=dH_n(Y;R)+[dH_n(Y;R)-\text{trace}\,T_{*n}](p-1)$$

so that $dH_n(X;R)+(p-1)\,\text{trace}\,T_{*n}=pdH_n(Y;R)$. Taking the alternating sum, we get $\chi(X;R)+(p-1)\alpha(T;R)=p\chi(Y;R)$. Using (2.4) and comparing with (2.6), we get $\alpha(T;R)=\chi(L;I_p)$.

Suppose $a>1$ and suppose the theorem has been proven for $a-1$. Consider $T_0=T^{q^{a-1}}$. Let Y_1 denote the orbit space of the map T_0 on X, and $f\colon X\to Y_1$ the natural decomposition map. Define a map $S\colon Y_1\to Y_1$ by $Sf=fT$. Then S is of period q^{a-1} on Y_1. Also, by (2.3), Y_1 has finitely generated integral cohomology groups. Hence, by the induction hypothesis, $\alpha(S;R)=\chi(L';I_q)$, where L' is the fixed point set of S.

We point out that L and L' are homeomorphic. Clearly, $f(L)\subset L'$ and f is 1-1 on L. We prove that $f(L)=L'$. Let $y\in L'$, where $y=f(x)$, $x\in X$. Then $f(x)=Sf(x)=fT(x)$ so $T(x)=T_0{}^k(x)$ for some k. But then $kq^{a-1}-1$ is a period for x, so $kq^{a-1}-1$ divides q^a. Hence $k=0$, so that $Tx=x$, and $x\in L$. So $\chi(L';I_q)=\chi(L;I_q)$.

Finally, $\alpha(S;R)=\alpha(T;R)$. Let $F_n=[x;x\in H_n(X;R),T_{0*}x=x]$. Then, by (2.3), f_* maps F_n isomorphically onto $H_n(Y_1;R)$. Moreover, since $S_*f_*=f_*T_*$, we have trace $(S_{*n};H_n(Y_1;R))=\text{trace}(T_*,F_n)$. But by (2.2), $\text{trace}(T_*,F_n)=\text{trace}(T_*,H_n(X;R))$. It follows that $\alpha(S;R)=\alpha(T;R)$ and the theorem follows.

We now turn to some results concerned with properties of periodic maps of large period.

(2.8) (*Smith*). *Let V be a finite dimensional rational vector space. There exists a positive integer r associated with V so that if T is any linear transformation on V with $T^p=$ identity where $p>r$, then there exists $1\leq j<p$ with $T^j=$ identity.*

Proof. We shall outline the proof ([9], pp. 161-162). Suppose $p = p_1^{a_1} p_2^{a_2} \cdots p_e^{a_e}$, where the p_i's are primes with $p_1 < p_2 < \cdots < p_e$. Define

$$\Phi(p) = \sum_1^e \phi(p_i^{a_i}) \text{ if } p_1 \neq 2 \text{ or } a_1 \neq 1; \; \Phi(p) = \sum_2^e \phi(p_i^{a_i}) \text{ otherwise.}$$

Then $\Phi(p) \to \infty$ as $p \to \infty$. We point out that if $\Phi(p) > dV$, then there exists $1 \leq j < r$ with $T^j =$ identity. For suppose this is not the case. Using the notation preceding (2.2), we have $f(t) = f_{s_1}(t) \cdots f_{s_k}(t)$, where $s_i \mid p$. Now each $p_i^{a_i}$ divides some s_j. For if not, each s_j divides $p/p_i = q$, so that $T^q =$ identity. But if each $p_i^{a_i}$ divides some s_j, it may be checked that $dV = \Sigma\, \phi(s_i) \geq \Phi(p)$. Hence $\Phi(p) \leq dV$, and the assertion follows.

(2.9) *As a consequence of* (2.8), *let X be a compact Hausdorff space with each $H_n(X; R)$ of finite dimension and $= 0$ for all but a finite number of n's. There exists a positive integer r so that if T is any periodic map on X, then $T_*{}^j : H_n(X; R) \to H_n(X; R)$ is, for some $1 \leq j \leq r$, the identity for all n.*

We denote the least such r by $r(X)$.

(2.10) THEOREM. *Let X be a finite dimensional compact Hausdorff space with finitely generated integral cohomology groups. Let T be a periodic map on X of period $p > r(X)$. There exists $1 \leq i < p$ such that $p/i = q$ is a prime, and such that if L_i denotes the fixed point set of T^i, then $\chi(X; R) = \chi(L_i, I_q)$.*

Proof. There exists, by (2.9), $1 \leq j \leq r$ with $T_{*n}{}^j =$ identity for all n. Suppose $p = j \cdot k \cdot q$, where k and q are positive integers with q prime. Let $i = j \cdot k$. Then $T_{*n}{}^i =$ identity for all n. Hence by (2.7), $\alpha(T^i; R) = \chi(X; R) = \chi(L_i, I_q)$.

The following is an extension of a result of Smith [9; 162]. It also generalizes the well-known theorem [11] that the periodic maps on a compact 2-manifold with negative Euler characteristic have uniformly bounded periods. It does not, however, provide the upper bound known for that case.

(2.11) THEOREM. *Let X be a compact manifold with $\chi(X; R) < 0$. Suppose T is a periodic map on X of period p, and such that if $1 \leq j < p$, then the dimension of the fixed point set of T^j is ≤ 1. Then $p \leq r(X)$.*

Proof. Suppose $p > r(X)$. Let i be the number given by (2.10). Then $\chi(X; R) = \chi(L_i; I_q) < 0$. But $\dim L_i \leq 1$, so that by a result of

Smith ([10], p. 704), L_i is the union of a disjoint collection of points and simple closed curves. Hence $\chi(L_i; I_q) \geq 0$, which is a contradiction.

(2.12) *The above theorem is not true if the restriction on the dimension of the fixed point set of T^j is removed.*

As an example, let X be a 2-sphere, and let Y be a 2-manifold with $\chi(Y) < 0$. Then $\chi(X \times Y) = \chi(X)\chi(Y) < 0$. But since X admits transformations of arbitrary period, so does $X \times Y$.

3. Convergence properties. We begin section 3 by stating an important result due to Smith [7] which is the basis for the work of this section. The result is stated and proved in the proof of Theorems I, II in [7].

(3.1) (*Smith*). *Let X be a locally compact n-dimensional Hausdorff space, $n < \infty$, and let T be a periodic map on X of prime period p. Denote by L the fixed point set of T. Suppose $0 \neq A_0 \subset A_1 \subset \cdots \subset A_m$, $m = pn + p$, is a sequence of compact subsets of X, with $T(A_i) = A_i$, and with every Čech cycle in A_i over I_p bounding in A_{i+1}. Then $L \cap A_m \neq 0$ and every cycle in $L \cap A_0$ over I_p bounds in $L \cap A_m$.*

We use also the concept of regular convergence introduced by Whyburn [12]. We shall phrase the definition in terms of Čech theory instead of Vietoris theory; these are interchangeable, as follows from the full equivalence of the two theories ([6], p. 277). Let X be a locally compact metric space, and let G be an abelian group. Let $[A_i]$ be a sequence of closed subsets of X, with A_i converging to a closed subset A of X. If n is a non-negative integer, then A_i converges n-regularly to A over G if and only if given $x \,\varepsilon\, A$ and a compact neighborhood U of x in X, there exists a closed neighborhood V of x (in X) with $V \subset U$, and a positive integer I, so that every Čech cycle in $V \cap A_i$ over G of dimension $\leq n$ bounds in $U \cap A_i$ for $i > I$. It may be seen that X is lc^n (i.e., homologically locally connected over G in the dimensions from 0 to n), if and only if the sequence $X, X, \cdots$ converges n-regularly to X.

Let X and Y be metric spaces. Let A_i be a sequence of closed subsets of X which converges to a subset A of X. Let $f_i: A_i \to Y$, $f: A \to Y$ be continuous. We shall say that f_i *converges continuously* to f if and only if whenever $x_i \to x$, $x_i \,\varepsilon\, A_i$, then $f_i(x_i) \to f(x)$. This specializes, in case $A_i = A$, to the notion of continuous convergence introduced by Carathéodory ([2], p. 58).

(3. 2) THEOREM. *Let X be a locally compact n-dimensional metric space. $n < \infty$. Suppose $[A_i]$ is a sequence of closed subsets of X converging n-regularly over I_p, p prime, to the subset A of X. Let T_i be a continuous periodic transformation of period p on A_i, such that $[T_i]$ converges continuously to the continuous function T on A. Then the fixed point set $[F_i]$ of T_i converges n-regularly over I_p to the fixed point set F of T.*

Proof. The reader may verify that if $x \varepsilon F$ and U is a neighborhood of x (in X), then there exists a neighborhood V of x and a positive integer I such that if $i > I$, then $\bigcup_j T_i{}^j(V \cap A_i) \subset U$.

Let $x \varepsilon F$ and let U be a compact neighborhood of x. There exists a sequence $U = U_{2m+1} \supset U_{2m} \supset \cdots \supset U_0$, $m = pn + p$, of compact neighborhoods of x (in X) and a positive integer I, such that $U_0 \cap A_i \neq 0$, for $i > I$, and (a) if $i > I$, then $\bigcup_j T_i{}^j(U_k \cap A_i) \subset U_{k+1}$ for $k = 0, \cdots, 2m$, and (b) for $i > I$ every cycle in $U_r \cap A_i$ over I_p bounds in $U_{r+1} \cap A_i$.

For each $0 \leq k \leq m$ and each $i > I$, define $V_{k,i} = \bigcup_j T_i{}^j(U_{2k} \cap A_i)$. Then $V_{k,i} \subset U_{2k+1} \cap A_i$, and $T_i(V_{k,i}) = V_{k,i}$. Moreover, since $V_{k+1,i} \supset U_{2k+2} \cap A_i$, every cycle in $V_{k,i}$ bounds in $V_{k+1,i}$. Hence we may apply (3. 1) to the sequence $V_{0,i} \subset V_{1,i} \subset \cdots \subset V_{m,i}$, and the transformation T_i. It follows that $V_{m,i} \cap F_i \neq 0$, and every cycle in $V_{0,i} \cap F_i$ bounds in $V_{m,i} \cap F_i$. Hence, for $i > I$ every cycle in $U_0 \cap F_i$ bounds in $U \cap F_i$, and $U \cap F_i \neq 0$.

To finish the proof, the reader has only to note that if $x_{m_i} \varepsilon F_{m_i}$, and $x_{m_i} \to x$, then $x \varepsilon F$. This follows easily from continuous convergence.

(3. 3) COROLLARY. *Let X be a locally compact n-dimensional metric space, $n < \infty$, which is lc^n over I_p, p prime. Let $[T_j]$ be a sequence of periodic maps on X of common period p^a, which converges continuously to the continuous map T. Then the fixed point set F_j of T_i converges n-regularly over I_p to the fixed point set F of T.*

Proof. The proof is a straight-forward combination of (3. 2) together with a procedure used often by Smith for extending proofs from period p to period p^a ([8], p. 367).

(3. 4) COROLLARY. *Let the hypotheses be those of* (3. 3) *and suppose in addition that X is compact. Then there exists I such that for $i > I$, we have $H_j(F_i; I_p) \approx H_j(F; I_p)$ for all j. In particular, suppose X an n-sphere. Then there exists an integer r so that F_i, $i > I$, and F are all homological r-spheres over I_p.*

Proof. This follows from a theorem of Begle [1].

(3.5) COROLLARY. *Let X be an n-dimensional compact metric space, $n < \infty$, which is lc^n over I_p, p prime. Let T be a periodic map on X of period p^a with fixed point set L. There is an $\epsilon > 0$ such that if T_1 is periodic on X of period p^a, $\rho(T(x), T_1(x)) < \epsilon$ for all $x \varepsilon X$, and L_1 denotes the fixed point set of T_1, then $H_j(L; I_p) \approx H_j(L_1; I_p)$ for all j.*

UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY.

[1] E. G. Begle, "Regular convergence," *Duke Mathematical Journal*, vol. 11 (1944), pp. 441-450.

[2] C. Carathéodory, *Conformal representation*, Cambridge, 1932.

[3] E. E. Floyd, "Examples of fixed point sets of periodic maps," *Annals of Mathematics*, vol. 55 (1952), pp. 167-171.

[4] ———, "On periodic maps and the Euler characteristics of associated spaces," *Transactions of the American Mathematical Society*, vol. 72 (1952), pp. 138-147.

[5] S. D. Liao, "A theorem on periodic transformations of homology spheres," *Annals of Mathematics* (to appear).

[6] S. Lefschetz, *Algebraic topology*, New York, 1942.

[7] P. A. Smith, "Fixed point theorems for periodic maps," *American Journal of Mathematics*, vol. 63 (1941), pp. 1-8.

[8] ———, "Fixed points of periodic transformations," Appendix B in [6].

[9] ———, "Periodic and nearly periodic transformations," *Lectures in Topology*, Ann Arbor, 1941, pp. 159-190.

[10] ———, "Transformations of finite period II," *Annals of Mathematics*, vol. 39 (1938), pp. 127-164.

[11] F. Steiger, "Die maximalen Ordnungen periodischer topologischer Abbildungen geschlosser Flächen in sich," *Commentarii Mathematici Helvetici*, vol. 8 (1935), pp. 48-69.

[12] G. T. Whyburn, "On sequences and limiting sets," *Fundamenta Mathematicae*, vol. 25 (1935), pp. 408-426.

TOPOLOGY OF METRIC COMPLEXES.*

By C. H. Dowker.

The metric complexes (polyhedra) discussed in this paper are metric spaces with a cell decomposition and an affine structure for each cell. These complexes are subject to certain mild conditions (section 9, conditions a and b′) which, for example, ensure local connectedness. The complexes are not, however, required to be finite or countable. They may be curved and they need not be locally finite.

If a complex is star-finite, and if the closed cells are given their usual topology, then the topology of the whole complex is uniquely determined. However, if the complex is not star-finite, there is no longer a unique topolgy. J. H. C. Whitehead has chosen for his topological polyhedra ([11], pages 315-321) the finest topology consistent with the usual topology for the closed cells. This is a very convenient and useful topology, but with it all complexes except star-finiteones become non-metrizable spaces.

In connection with his studies of local connectedness, S. Lefschetz ([9], Chapter I) has chosen two particular ways of giving a complex a metric. If the complex is not star-finite, the topology induced by each of these metrics is necessarily less fine than the Whitehead topology. In general, the two different Lefschetz metrics induce different topologies.

In this paper, instead of choosing some particular metric, we take a somewhat axiomatic point of view and state conditions which should be satisfied by any metric complex. Our theorems are then shown to be consequences of these conditions. However, our method is not one of proof directly from the axioms. Instead, we use the method of comparing each metric complex with the corresponding Whitehead complex, that is, the same complex retopologized with the Whitehead topology.

In the first chapter, we discuss affine complexes. These are sets which have a cell decomposition and an affine structure for each cell, but they have no topology. These affine complexes have homology and cohomology groups, and the theorem on invariance under subdivision holds.

In the second chapter we add topology to the affine complex, and state conditions on the topology in order that the complex may be called a topological complex.

* Received August 10, 1950.

In the third chapter we discuss Whitehead complexes, that is, affine complexes with the Whitehead topology. Since the method of investigating metric complexes is that of comparison with the corresponding Whitehead complexes, we give a rather complete resumé of the known theorems on Whitehead complexes.

The fourth and main chapter contains the results on metric complexes. A metric complex is defined to be a topological complex whose topology is induced by a metric. Given a metric complex we construct (sections 10-13) a sequence of locally finite coverings of the complex by open sets, and using this sequence of coverings (sections 14-15) we construct a homotopy of the identity mapping of the complex. Then (section 16) by means of this homotopy, we prove that each metric complex has the same homotopy type as the corresponding Whitehead complex. It follows that any two isomorphic metric complexes have the same homotopy type. In section 17 we discuss the mappings of a space in the nerve of a covering when this nerve is a metric complex. In section 18 we show the topological invariance of the homology and cohomology groups of metric complexes.

I. Affine Complexes.

1. Definition and properties of affine complexes. By a convex cell E of a Euclidean space we mean an open bounded convex cell; its closure $\bar{E}$ is called a closed convex cell.

Let a set X be the union of a family $\{e_\alpha\}$ of mutually disjoint subsets e_α of K. Let each e_α be associated with a 1-1 transformation ϕ_α of some closed convex cell $\bar{E}_\alpha$ into X such that (i) ϕ_α maps the convex cell E_α onto e_α and (ii) if E' is any face of E, then $\phi_\alpha(E')$ is an e_β, and $\phi_\alpha^{-1}\phi_\beta$ is an affine (linear) transformation of $\bar{E}_\beta$ onto $\bar{E}'$. Then the set X, together with the decomposition $\{e_\alpha\}$ and the family $\{\phi_\alpha\}$ of transformations, is called an *affine complex.*

Each of the subsets e_α with the linear structure given by the transformation $\phi_\alpha \mid E_\alpha : E_\alpha \to e_\alpha$, is called a cell of the complex. The dimension of the cell e_α is defined to be the dimension of E_α. Each cell of dimension zero contains a single point which is called a vertex. If $e_\beta = \phi_\alpha(E')$, where E' is a face of E_α, e_β is called a face of e_α; we write $e_\beta \leq e_\alpha$. If E' is a proper face of E_α, e_β is called a proper face of e_α; we write $e_\beta < e_\alpha$. By a *finite* affine complex we mean one with only a finite number of cells. By a *star-finite* affine complex we mean an affine complex which, for each cell e_β, has only a finite number of cells e_α with $e_\beta \leq e_\alpha$.

If K is an affine complex, the cells e_α of K may be oriented by assigning orientations to the corresponding convex cells E_α. If e_β is an $(r-1)$-dimensional face of an r-cell e_α, the incidence number $[e_\alpha : e_\beta]$ is defined to be the incidence number $[E_\alpha : E']$, provided $\phi_\alpha^{-1}\phi_\beta : \bar{E}_\beta \to \bar{E}'$ is orentation preserving, and to be $-[E_\alpha : E']$ otherwise. With such a definition of orientation and incidence numbers, the affine complex becomes a closure finite oriented cell complex ([8], page 89). Thus, given a topological abelian group G and a non-negative integer p, we may define the p-dimensional cohomology group $H^p(K, G)$, and if G is discrete we may define the p-dimensional homology group $H_p(K, G)$.

The elements of the underlying set X of an affine complex K are called the points of K. A closed cell $\bar{e}_\alpha$ of K (the *closure* of e_α) is defined to be the image set $\phi_\alpha(\bar{E}_\alpha)$ together with the transformation ϕ_α. If x and y are points of $\bar{e}_\alpha$, the closed segment $[x, y]$ of $\bar{e}_\alpha$ is defined to be the image by ϕ_α of the closed segment of $\bar{E}_\alpha$ joining $\phi_\alpha^{-1}(x)$ to $\phi_\alpha^{-1}(y)$. If $0 \leq t \leq 1$, the point $tx + (1-t)y$ which divides the segment from x to y in $\bar{e}_\alpha$ in the ratio $t : 1-t$ is defined to be the image by ϕ_α of the point dividing the segment from $\phi_\alpha^{-1}(x)$ to $\phi_\alpha^{-1}(y)$ in the ratio $t : 1-t$. Similarly, a convex set in $\bar{e}_\alpha$ is defined to be the image by ϕ_α of a convex set of $\bar{E}_\alpha$. If A is a subset of $\bar{e}_\alpha$, the convex hull A^* of A in $\bar{e}_\alpha$ is the image by ϕ_α of the convex hull of $\phi_\alpha^{-1}A$ in $\bar{E}_\alpha$; thus A^* is the least convex set of $\bar{e}_\alpha$ containing A.

Note that $[x, y]$ may depend on $\bar{e}_\alpha$ as well as on x and y. However, if e_β is a face of e_α, and if x and y are points of $\bar{e}_\beta$, then the segment $[x, y]$ in $\bar{e}_\beta$ is the same as the segment $[x, y]$ in $\bar{e}_\alpha$, and the point $tx + (1-t)y$ in $\bar{e}_\beta$ is the same as the point $tx + (1-t)y$ in $\bar{e}_\beta$. In fact, since $\phi_\alpha^{-1}\phi_\beta$: $\bar{E}_\beta \to \bar{E}_\alpha$ is affine, the affine structure of $\bar{e}_\beta$ is that induced by the affine structure of $\bar{e}_\alpha$.

It is clear that the set $\bar{e}_\alpha$ is the union of all the faces of e_α; $\bar{e}_\alpha = \bigcup_{e_\beta \leq e_\alpha} e_\beta$. The star $\operatorname{St} e_\alpha$ of a cell e_α of K is defined to be the union of all cells e_β such that e_α is a face of e_β; $\operatorname{St} e_\alpha = \bigcup_{e_\alpha \leq e_\beta} e_\beta$. Note that $e_\alpha \subset \bar{e}_\alpha$, $e_\alpha \subset \operatorname{St} e_\alpha$, $e_\alpha = \bar{e}_\alpha \cap \operatorname{St} e_\alpha$. The following three statements are equivalent: (i) $e_\beta \leq e_\alpha$, (ii) $\bar{e}_\beta \subset \bar{e}_\alpha$, (iii) $\operatorname{St} e_\alpha \subset \operatorname{St} e_\beta$.

If x is a point of K, $e(x)$ is defined to be the unique cell e_α containing x, and $\bar{e}(x)$ is defined to be the closure of $e(x)$. The following five statements are equivalent: (i) $x \in \bar{e}(y)$, (ii) $\bar{e}(x) \subset \bar{e}(y)$, (iii) $e(x) \leq e(y)$, (iv) $\operatorname{St} e(y) \subset \operatorname{St} e(x)$, (v) $y \in \operatorname{St} e(x)$.

Two affine complexes K_1 and K_2 are called *isomorphic* if there is a 1-1 order preserving correspondence between the set of cells of K_1 and the set of

cells of K_2. The correspondence is called an isomorphism. It can be shown (See [1], page 127) that under an isomorphism, each cell corresponds to a cell of the same dimension.

Suppose there is given a subcollection of cells of an affine complex K such that if any cell e_α is in the subcollection, so is each face of e_α. Then the union L of the cells e_α of the subcollection is a set of points decomposed into cells e_α which are associated with transformations $\phi_\alpha : \bar{E}_\alpha \rightarrow L$. In fact L is an affine complex which is called a *subcomplex* of K. In particular, if e_α is a cell of K, the set $\bar{e}_\alpha$ with the obvious cell decomposition is a subcomplex of K. Also, the union of the closures of the cells contained in $\mathrm{St}\, e_\alpha$, with the obvious cell decomposition, forms a subcomplex which we call $\overline{\mathrm{St}}\, e_\alpha$.

If K and L are affine complexes, the product set $K \times L$ can be decomposed into cells $e_{\alpha\beta} = e_\alpha \times e_\beta$. For each $e_{\alpha\beta}$ let $\bar{E}_\alpha \times \bar{E}_\beta$ be the closed convex cell which is the cartesian product of $\bar{E}_\alpha$ and $\bar{E}_\beta$, and let $\phi_{\alpha\beta} : \bar{E}_\alpha \times \bar{E}_\beta \rightarrow K \times L$ be defined by $\phi_{\alpha\beta}(x, y) = (\phi_\alpha(x), \phi_\beta(y))$. It is easily verified that $K \times L$ thus becomes an affine complex. This complex is called the *product complex* of K and L.

An affine complex K is called *simplicial* if (i) for each e_α, E_α is a simplex, and (ii) each non-empty intersection $\bar{e}_\alpha \cap \bar{e}_\beta$ of two closed cells of K is a closed cell $\bar{e}_\gamma$. For a discussion of simplicial affine complexes see ([9], § 4).

2. Subdivision. A *subdivision* of an affine complex K is a 1-1 transformation $\mathrm{Sd} : K \rightarrow K'$ of K onto an affine complex K' such that (i) the image of each cell of K consists of the union of a finite number of cells, and (ii) the inverse transformation is linear on each closed cell of K'. We shall also say that the affine complex K' is a subdivision of K.

By condition (i), for each cell e_α of K' there is a unique cell e_β of K such that $e_\alpha \subset \mathrm{Sd}\, e_\beta$. Condition (ii) means that, if e_α and e_β have affine structures given by $\phi_\alpha : \bar{E}_\alpha \rightarrow \bar{e}_\alpha$ and $\phi_\beta : \bar{E}_\beta \rightarrow \bar{e}_\beta$, then $\phi_\beta^{-1}\mathrm{Sd}^{-1}\phi_\alpha : \bar{E}_\alpha \rightarrow \bar{E}_\beta$ is linear.

(2. 1) *Isomorphic affine complexes have isomorphic simplicial subdivisions.*

Proof. The barycentric subdivision ([1], page 135) of an affine complex is a simplicial affine complex. Isomorphic affine complexes have isomorphic barycentric subdivisions.

(2. 2) *If K and L are isomorphic simplicial affine complexes, there*

exists a map f of K onto L which maps each cell of K onto a corresponding cell of L, and which is linear on each closed cell.

Proof. The natural barycentric mapping ([9], p. 7; [1], p. 138, § 6) has the required properties.

Given a subdivision $\mathrm{Sd}: K \to K'$ of an affine complex K, let a chain transformation $\rho_p: C_p(K) \to C_p(K')$ be defined as follows: If $e_\beta{}^p$ is an elementary p-chain of K, let $\rho_p e_\beta{}^p = \Sigma \epsilon_\alpha e_\alpha{}^p$, where the summation is over all α such that $e_\alpha{}^p$ is a p-cell of K', $e_\alpha{}^p \subset \mathrm{Sd}\, e_\beta{}^p$, and $\epsilon_\alpha = 1$ if $\phi_\beta{}^{-1}\mathrm{Sd}^{-1}\phi_\alpha$ is orientation preserving, $\epsilon_\alpha = -1$ if it is orientation reversing.

It is assumed known that, in the finite complex $\mathrm{Sd}\, \bar{e}_\beta$, $\rho\partial e_\beta{}^p = \partial\rho e_\beta{}^p$. Hence $\rho\partial = \partial\rho: C_p(K) \to C_{p-1}(\mathrm{K}')$; thus ρ is a chain mapping ([6], p. 411; [8], p. 145). It is to be shown that ρ is a chain equivalence ([6], page 414).

If $e_\alpha \subset \mathrm{Sd}\, e_\beta$, let $\mathrm{T}\, e_\alpha$ be the finite subcomplex $\bar{e}_\beta$ of K. We define chain transformations $\tau_p: C_p(K') \to C_p(K)$ such that (i) $\tau_{p-1}\partial = \partial\tau_p: C_p(K') \to C_{p-1}(K)$, and (ii) $\tau_p e_\alpha{}^p$ is a chain of the subcomplex $\mathrm{T}\, e_\alpha{}^p$ of K. For each elementary 0-chain $e_\alpha{}^0$, $\tau\, e_\alpha{}^0$ is chosen to be any elementary 0-chain in $\mathrm{T}\, e_\alpha{}^0$. If $e_\alpha{}^1$ is an elementary 1-chain, $\tau\partial\, e_\alpha{}^1$ is a bounding 0-cycle in $\mathrm{T}\, e_\alpha{}^1$, and we chose as $\tau\, e_\alpha{}^1$ any 1-chain in $\mathrm{T}\, e_\alpha{}^1$ bounded by $\tau\partial\, e_\alpha{}^1$. If $p > 1$, assume that τ has been defined for dimensions less than p. Then $\tau\partial\, e_\alpha{}^p$ has been defined and is a chain in $\mathrm{T}\, e_\alpha{}^p$. Since $\partial\tau\partial\, e_\alpha{}^p = \tau\partial\partial\, e_\alpha{}^p = 0$, $\tau\partial\, e_\alpha{}^p$ is a cycle in $\mathrm{T}\, e_\alpha{}^p$; hence, since $\mathrm{T}\, e_\alpha{}^p$ is acyclic in dimension greater than zero, $\tau\partial\, e_\alpha{}^p$ is a bounding cycle in $\mathrm{T}\, e_\alpha{}^p$. Let $\tau\, e_\alpha{}^p$ be chosen as a chain of $\mathrm{T}\, e_\alpha{}^p$ whose boundary is $\tau\partial\, e_\alpha{}^p$.

In particular, if K and K' are simplicial, there is a simplicial map $\pi: K' \to \mathrm{K}$, called a projection, which maps each vertex v of K' into a vertex of $\mathrm{T}\, v$. Then π induces chain mappings $\pi_p: C\ (K') \to C_p(K)$ such that $\pi_p\, e_\alpha{}^p$ is a chain of $\mathrm{T}\, e_\alpha{}^p$. We may then take $\tau_p = \pi_p$.

It is to be shown that $\tau_p\rho_p: C_p(K) \to C_p(K)$ is the identity. This is clear for $p = 0$. Assume that $p > 0$ and that it is proved for dimensions less than p. Then $\tau\rho\, e^p$ is a chain of $\mathrm{TSd}\, e_\alpha{}^p = \bar{e}_\alpha{}^p$, and $\partial\tau\rho\, e_\alpha{}^p = \tau\rho\partial\, e_\alpha{}^p = \partial\, e_\alpha{}^p$ by the induction hypothesis. But the only chain of $\bar{e}_\alpha{}^p$ with boundary $\partial\, e_\alpha{}^p$ is the chain $e_\alpha{}^p$. Hence $\tau\rho\, e_\alpha{}^p = e_\alpha{}^p$. It follows that $\tau\rho$ is the identity chain mapping.

We now show that the chain mapping $\rho\tau$ is chain homotopic to the identity. We define a homomorphism $D_{p+1}: C_p(K') \to C_{p+1}(K')$ so that $\partial Dc^p + D\partial c^p = c^p - \rho\tau c^p$, and so that $De_\alpha{}^p$ is a chain of $\mathrm{SdT}\, e_\alpha{}^p$. If $e_\alpha{}^0$ is an elementary 0-chain, $e_\alpha{}^0 - \rho\tau e_\alpha{}^0 - D\partial e_\alpha{}^0 = e_\alpha{}^0 - \rho\tau e_\alpha{}^0$ is a bounding 0-chain in $\mathrm{SdT}\, e_\alpha{}^0$. Let $D_1 e_\alpha{}^0$ be chosen as a 1-chain in $\mathrm{SdT}\, e_\alpha{}^0$ whose boundary is $e_\alpha{}^0 - \rho\tau_\alpha{}^0$. Suppose that, for $p > 0$, D has been defined for chains of

dimension less than p. Then $D\partial e_a{}^p$ has been defined. Let $c_p = e_a{}^p - \rho\tau e_a{}^p - D\partial e_a{}^p$. Then c_p is a p-chain in SdT $e_a{}^p$, and by computation one finds that $\partial c_p = 0$. Thus c_p is a p-cycle in SdT $e_a{}^p$, a subcomplex which is acyclic in dimensions greater than zero. Hence c_p is a bounding cycle in SdT $e_a{}^p$. Let $D_{p+1}e_a{}^p$ be chosen as a $(p+1)$-chain in SdT $e_a{}^p$ bounded by c_p.

Thus $\tau\rho$ is the identity and $\rho\tau$ is chain homotopic to the identity. Therefore, ρ is a chain equivalence, and we have

(2.3) *If K' is a subdivision of an affine complex K, the homology group $H_p(K')$ is isomorphic with $H_p(K)$, and the cohomology group $H^p(K')$ is isomorphic with $H^p(K)$. If K and K' are simplicial, the homomorphisms $\pi_*: H_p(K') \to H_p(K)$ and $\pi^*: H^p(K) \to H^p(K')$ induced by the projection $\pi: K' \to K$ are isomorphisms onto.*

II. Topological complexes.

3. Definition of topological complexes. An affine complex K is called a topological complex if its underlying point set is a topological space, and if

(a) each $\phi_a: \bar{E}_a \to \bar{e}_a$ is a homeomorphism,

(b) for each neighborhood U of each point x of K, there is some neighborhood V of x such that, for each point y in V, $x \in \bar{e}(y)$, and the segment $[x, y]$ in $\bar{e}(y)$ is contained in U.

From condition (b) it follows immediately that as a space, K is locally connected. Also by condition (b), each point x has a neighborhood V such that, for each $y \in V$, $y \in \operatorname{St} e(x)$, hence such that $V \subset \operatorname{St} e(x)$. If e_a is any cell of K, and if $x \in \operatorname{St} e_a$, then $e_a \leq e(x)$, and $\operatorname{St} e(x) \subset \operatorname{St} e_a$. Thus each point x of $\operatorname{St} e_a$ has a neighborhood $V \subset \operatorname{St} e_a$; hence $\operatorname{St} e_a$ is open. Thus the star of any cell is an open set containing the cell. However, as the example below shows, one can have open stars and local connectedness with condition (b) not satisfied.

The complement in K of a closed cell $\bar{e}_a$ is the union of the stars of the cells in the complement; hence $K - \bar{e}_a$ is open, and $\bar{e}_a$ is a closed set in K. In fact, $\bar{e}_a$ is the topological closure in K of the subset e_a. By condition (a), $\bar{e}_a$ is compact.

Example. Let X be the subset of the cartesian plane with the following cell decomposition. The 0-cells of K are $A_0:(1,0)$ and $A_n:(2,1/n)$ for $n = 1, 2, \cdots$. The 1-cells of K are the segments A_nA_{n+1} and the broken lines A_0, $(-1/n, -1/n)$, $(-1/n, 1/n)$, A_n. The 2-cells of K are the regions

$A_0A_nA_{n+1}$ bounded by the 1-cells A_nA_{n+1}, A_0A_n and A_0A_{n+1}. Some suitable choice of the maps ϕ_α is to be made.

It may easily be verified that the space X of the example is locally connected, and that the star of each cell is an open set. However, condition (b) is not satisfied at the point A_0, and hence the complex is not a topological complex.

If K is any topological complex, an affine subcomplex is a topological subspace. Clearly conditions (a) and (b) hold also for the subcomplex. Thus a subcomplex of a topological complex is a topological complex.

If K_1 and K_2 are topological complexes, then $K_1 \times K_2$ is an affine complex and is also a topological space, the topological product of K_1 and K_2. Clearly $\phi_{\alpha\beta}: \bar{E}_\alpha \times \bar{E}_\beta \to \bar{e}_{\alpha\beta}$ $(= \bar{e}_\alpha \times \bar{e}_\beta)$ is a homeomorphism if $\phi_\alpha: \bar{E}_\alpha \to \bar{e}_\alpha$ and $\phi_\beta: \bar{E}_\beta \to \bar{e}_\beta$ are homeomorphisms. Thus condition (a) is satisfied. Any neighborhood of a point (x_1, x_2) of $K_1 \times K_2$ contains a neighborhood of the form $U_1 \times U_2$. Let neighborhoods V_1 of x_1 in K_1 and V_2 of x_2 in K_2 be chosen as in condition (b). Then $V_1 \times V_2$ is a neighborhood of (x_1, x_2) such that if $(y_1, y_2) \in V_1 \times V_2$, then $x_1 \in \bar{e}(y_1)$, $x_2 \in \bar{e}(y_2)$, and hence $(x_1, x_2) \in \bar{e}(y_1) \times \bar{e}(y_2) = \bar{e}(y_1, y_2)$, and the segment $[(x_1, x_2), (y_1, y_2)]$ in $\bar{e}(y_1, y_2)$ is contained in $[x_1, y_1] \times [x_2, y_2] \subset U_1 \times U_2$. Thus condition (b) is satisfied, and we have

(3.1) *the product of two topological complexes is a topological complex.*

III. Whitehead complexes.

4. The Whitehead topology. An arbitrary affine complex K can be made into a topological complex by giving it the finest[1] ([2], p. 9) topology consistent with condition (a). That is (cf. [11], p. 316), a set of K is called open if and only if the intersection with each closed cell $\bar{e}_\alpha$ is the image by ϕ_α of an open set of $\bar{E}_\alpha$. If U is any open set containing a given point x of K, let V be the set of points y of $\mathrm{St}\, e(x)$ such that the segment $[x, y]$ in $\bar{e}(y)$ is contained in U. It can be seen that the intersection of V with each closed cell $\bar{e}_\alpha$ is an open set of $\bar{e}_\alpha$; hence V is open. Clearly $x \in V$. Thus condition (b) is satisfied. This finest topology will be called the Whitehead topology, and the resulting topological complex will be called a Whitehead complex. It is known ([11], p. 320; [12], p. 225) that a Whitehead complex is a normal Hausdorff space.

---

[1] "Fine" in the sense of Bourbaki means "weak" in the sense of Whitehead or "strong" as used in functional analysis.

An equivalent description of the Whitehead topology is that a set of K is called closed if and only if its intersection with each closed cell $\bar{e}_a$ is the image by ϕ_a of a closed set of E.

It can be shown ([11], pp. 316-317) that if K is star-finite (in particular, finite), the Whitehead topology is the only one which satisfies conditions (a) and (b). Accordingly, given a star finite affine complex K, one may speak unambiguously of an open (or closed) set of K.

(4.1) *If L is a subcomplex of a Whitehead complex K, the subspace topology of L coincides with its Whitehead topology.*[2]

Proof. If A is a subset of L closed in the Whitehead topology of L, then $A \cap \bar{e}_a$ is closed in $\bar{e}_a$ for every $\bar{e}_a$ in L, hence also for every $\bar{e}_a$ in K. Therefore A is closed in K, and hence $A = A \cap L$ is closed in the subspace topology of L. Thus the Whitehead topology is not finer than the subspace topology, and the two topologies coincide.

(4.2) *A transformation f of a Whitehead complex K into a topological space Y is continuous if and only if $f|\bar{e}_a$ is continuous for each closed cell $\bar{e}_a$ of K.*

Proof.[3] If f is continuous, $f^{-1}(V)$ is open in K for each open set V of Y; hence $f^{-1}(V) \cap \bar{e}_a$ is open in $\bar{e}_a$, and $f|\bar{e}_a$ is continuous. Conversely, if $f|\bar{e}_a$ is continuous, $f^{-1}(V) \cap \bar{e}_a$ is open in $\bar{e}_a$, and hence $f^{-1}(V)$ is open in K; therefore f is continuous.

(4.3) *A transformation f of a Whitehead complex K into a topological space Y is continuous if and only if $f|L$ is continuous for each finite subcomplex L of K.*

Proof . If f is continuous, then $f|L$ is continuous for every subspace L, in particular for each finite subcomplex. If $f|L$ is continuous for each finite subcomplex L, then in particular $f|\bar{e}_a$ is continuous for each closed cell $\bar{e}_a$, and hence, by (4.2), f is continuous.

5. The product complex. If K is an affine complex, we denote by K_W the same complex with the Whitehead topology.

(5.1) *Let $K \times L$ be the product of two affine complexes. Then if either K or L is star-finite, $(K \times L)_W = K_W \times L_W$.*

---

[2] See Whitehead ([12], p. 224).

[3] See Whitehead ([11], p. 317; [12], p. 224).

Proof.[4] By (3.1), $K_W \times L_W$ is a topological complex. Hence the product topology is not finer than the Whitehead "finest" topology. It is then sufficient to show that if a set G is open in $(K \times L)_W$, it is open in $K_W \times L_W$; that is, for each point (x_0, y_0) of G there exist open sets U and V of K_W and L_W respectively, such that $x_0 \varepsilon U$, $y_0 \varepsilon V$ and $U \times V \subset G$.

Assume that it is L which is star-finite. For any cell $\bar{e}_\beta$ of L, $\bar{e}(x_0) \times \bar{e}_\beta = (\bar{e}(x_0) \times \bar{e}_\beta)_W$ is a finite subcomplex of $(K \times L)_W$, and $G_\beta = G \cap (\bar{e}(x_0) \times \bar{e}_\beta)_W$ is open in $\bar{e}(x_0) \times \bar{e}_\beta$. Hence $\{y \mid (x_0, y) \varepsilon G_\beta\}$ is open in $\bar{e}_\beta$. Let $H = \{y \mid (x_0, y) \varepsilon G\}$; then $H \cap \bar{e}_\beta = \{y \mid y \varepsilon \bar{e}_\beta, (x_0, y) \varepsilon G\}$ $= \{y \mid (x_0, y) \varepsilon G_\beta\}$, which is open in $\bar{e}_\beta$. Hence H is open in L_W.

Clearly $y_0 \varepsilon H$. Since L_W is a normal space, we can choose an open set V of L_W so that $y_0 \varepsilon V$, $\bar{V} \subset H$, and $V \subset \operatorname{St} e(y_0)$. Let $M = \overline{\operatorname{St}}\, e(y_0)$; since L is star-finite, M is a finite subcomplex of L. Then $y_0 \varepsilon V \subset \bar{V} \subset H \cap M$. Let the subset U of K be defined by $U = \{x \mid x \times \bar{V} \subset G\}$. Since $\bar{V} \subset H$, $x_0 \varepsilon U$. Hence $(x_0, y_0) \varepsilon U \times V \subset G$. Thus there remains to be shown only that U is open in K_W.

If e_a is a cell of K, $\bar{e}_a \times M = (\bar{e}_a \times M)_W$ is a finite subcomplex of $(K \times L)_W$. Hence $G_a = G \cap (\bar{e}_a \times M)_W$ is an open set of $\bar{e}_a \times M$. Then $U \cap \bar{e}_a = \{x \mid x \varepsilon \bar{e}_a, x \times \bar{V} \subset G\} = \{x \mid x \times \bar{V} \subset G_a\}$, which is open in $\bar{e}_a$, since G_a is open and $\bar{V}$ is compact. Hence U is open in the Whitehead complex K_W. This completes the proof of (5.1).

The condition of star-finiteness of one factor can not be dropped as the following example shows.

Example. Let K consist of a collection of closed 1-cells A_i of the power of the continuum, with a common vertex u_0. Let L be a countably infinite collection of closed 1-cells B_j, $j = 1, 2, \cdots$, with a common vertex v_0. Then $(K \times L)_W \neq K_W \times L_W$.

Proof. Let the closed 1-cells A_i have parameters x_i, $0 \leq x_i \leq 1$ so that, at the point u_0, $x_i = 0$ for all i. Let the closed 1-cells B_j have parameters y_j, $0 \leq y_j \leq 1$, and at v_0, let $y_j = 0$ for all j. Let the indices i be sequences of integers; $i = \{i_1, i_2, \cdots\}$. For each pair (i, j) of indices, let p_{ij} be the point $(1/i_j, 1/i_j)$ of $A_i \times B_j \subset K \times L$, and let P be the set of all such points p_{ij}. Then for each $A_i \times B_j$, $P \cap (A_i \times B_j)$ consists of one point p_{ij}, hence is closed in $A_i \times B_j$. Thus P is closed in $(K \times L)_W$.

A neighborhood U of u_0 in K_W is given by $x_i < a_i$, where the a_i are positive numbers. A neighborhood V of v_0 in L_W is given by $y_j < b_j$, $b_j > 0$.

[4] See Whitehead ([12], p. 227).

Let $U \times V$ be a product neighborhood of (u_0, v_0) in $K_W \times L_W$. Let $\bar{\imath} = \{\bar{\imath}_1, \bar{\imath}_2, \cdots\}$ be chosen so that for each j, $\bar{\imath}_j > j$ and $\bar{\imath}_j > (b_j)^{-1}$. Let $\bar{\jmath}$ be chosen so $\bar{\jmath} > (a_{\bar{\imath}})^{-1}$. Then $(\bar{\imath}_{\bar{\jmath}})^{-1} < (\bar{\jmath})^{-1} < a_{\bar{\imath}}$, and $(\bar{\imath}_{\bar{\jmath}})^{-1} < b_{\bar{\jmath}}$. Hence $p_{\bar{\imath}\bar{\jmath}} \in U \times V$. Therefore every neighborhood of (u_0, v_0) in $K_W \times L_W$ contains a point of P. Hence, since $(u_0, v_0) \notin P$, P is not closed in the product topology. Thus the product topology is not so fine as the Whitehead topology.[5]

(5.2) *Let K be a Whitehead complex, let I be the closed interval $0 \leq t \leq 1$, and let Y be a space. Let h be a function from $K \times I$ to Y such that, for each closed cell $\bar{e}_\alpha$ of K, $h \mid \bar{e}_\alpha \times I$ is continuous. Then $h: K \times I \to Y$ is continuous.*[6]

Proof. We regard I as a complex consisting of one 1-cell and two 0-cells. Then by (5.1), since I is a finite (and hence star-finite) complex, $K \times I$ is a Whitehead complex. But for each closed cell $\bar{e}_\beta$ of $K \times I$, there is a closed cell $\bar{e}_\alpha$ of K such that $\bar{e}_\beta \subset \bar{e}_\alpha \times I$. Therefore, since $h \mid \bar{e}_\alpha \times I$ is continuous, $h \mid \bar{e}_\beta$ is continuous. It follows from (4.2) that h is continuous.

6. Subdivisions of Whitehead complexes. One can show that any subdivision of a Whitehead complex is a Whitehead complex. More exactly, we have

(6.1) *If* Sd$: K \to K'$ *is a subdivision of an affine complex K, and if both K and K' are given the Whitehead topology, then* Sd *is a homeomorphism.*

Proof. Sd^{-1} maps each closed cell of K' linearly, and hence continuously, into a closed cell of K. Hence Sd^{-1} is continuous. Sd maps each finite subcomplex L of K onto a finite subcomplex L' of K', and Sd/L is piecewise linear, hence continuous. Thus Sd is continuous, and therefore a homeomorphism.

(6.2) *Isomorphic Whitehead complexes are homeomorphic.*

Proof. Let K and L is isomorphic Whitehead complexes. If K and L are simplicial, the natural barycentric map of K on L is linear and hence continuous on each closed cell, and its inverse has the same property. Hence the natural barycentric map is a homeomorphism. If K and L are not simplicial, then by (2.1) they have isomorphic, and hence homeomorphic, simplicial subdivisions K' and L'. But by (6.1) K is homeomorphic to K', and L to L'. Hence K and L are homeomorphic.

[5] This answers a question of Whitehead, see [12, page 227, footnote 21].

[6] See Whitehead ([12], p. 228).

7. Coverings. A covering of a space X is a collection $\{U_\alpha\}$ of open set whose union is X. A covering $\{U_\alpha\}$ is called locally finite if every point of X has a neighborhood which meets only a finite number of sets U_α of the covering. A covering $\{V_\beta\}$ is called a refinement of $\{U_\alpha\}$, if each V_β is contained in some U_α. If K is a topological complex, the star of each vertex v_α of K is an open set, and every point of K is in the star of some vertex; hence $\{\mathrm{St}\, v_\alpha\}$ is a covering of K.

(7.1) *If K is a Whitehead complex and if $\{U_\alpha\}$ is a covering of K. there exists a simplicial subdivision* $\mathrm{Sd}: K \to K'$ *such that the covering of K' by the stars of its vertices is a refinement of the covering* $\{\mathrm{Sd}\, U_\alpha\}$.

For the proof see Whitehead ([11], Theorem 35).

Let X be a topological space, let $\mathfrak{U}$ be a covering of X, and let N $(= N(\mathfrak{U}))$ be the nerve of $\mathfrak{U}$ topologized as a Whitehead complex. Then N is a simplicial complex with a vertex u_α corresponding to each non-empty set of $\mathfrak{U}$, and a simplex $u_{\alpha_0}, \cdots, u_{\alpha_p}$ corresponding to each non-empty intersection $U_{\alpha_0} \cap \cdots \cap U_{\alpha_p}$. A mapping $\phi\colon X \to N$ is called canonical with respect to $\mathfrak{U}$ if for each $U_\alpha \in \mathfrak{U}$, $\phi^{-1}\mathrm{St}\, u_\alpha \subset U_\alpha$.

(7.2) *If $\mathfrak{U}$ is a locally finite covering of a normal space X, there exists a canonical mapping of X into the Whitehead nerve N of $\mathfrak{U}$.*

Proof.[7] Let M be the nerve of $\mathfrak{U}$ metrized with the "natural" metric ([9], (4.12)). Then there exists ([4], 3(a); [3], Theorem 1.1) a canonical map θ of X into M. Now M and N are the same affine complex with two different topologies; let $\chi\colon M \to N$ be the identity map. Then χ is linear and hence continuous on each closed cell of M. Hence χ is continuous on each finite subcomplex of M. Hence by [3], Lemma 1.2, $\chi\theta\colon X \to N$ is continuous. Let $\phi = \chi\theta$; clearly ϕ is a canonical map of X into N.

8. Homology and cohomology groups. Let K be a Whitehead complex, and let $|K|$ be its underlying space. If K' is a subdivision of K, we identify the spaces $|K|$ and $|K'|$ by means of the homeomorphism $\mathrm{Sd}\colon |K| \to |K'|$. By the Čech cohomology groups of K we shall mean the Čech cohomology groups[8] of the space $|K|$.

(8.1) *The Čech cohomology groups of a Whitehead complex K are isomorphic with the corresponding combinatorial cohomology groups of K.*

---

[7] It may be seen that the proof of [3], Theorem 1.1 or the proof outlined for [4], proposition (a) does not depend on the special topology of the nerve; thus either of these gives a direct proof of (7.2).

[8] For definition and properties of Čech cohomology groups see [5].

Proof. Let K'_0 be a fixed simplicial subdivision of K. For each simplicial subdivision K'_α of K'_0, the nerve of the covering $\mathfrak{U}_\alpha$ of $|K|$ by the stars of the vertices of K'_α is a complex which can be identified with K'_α itself. By (7.1) these coverings $\mathfrak{U}_\alpha$ of $|K|$ by stars of vertices of simplicial subdivisions of K'_0 form a cofinal family of coverings. By (2.3), the projection map $\pi_{\alpha 0}$ of the nerve K'_α of $\mathfrak{U}_\alpha$ into the nerve K'_0 of $\mathfrak{U}_0$, induces an isomorphism $\pi^*_{\alpha 0}$ of $H^p(K'_0)$ onto $H^p(K'_\alpha)$. If K'_α and K'_β are two simplicial subdivisions of K'_0 such that $\mathfrak{U}_\beta$ is a refinement of $\mathfrak{U}_\alpha$, then ([5], p. 282) $\pi^*_{\beta 0} = \pi^*_{\beta\alpha}\pi^*_{\alpha 0}: H^p(K'_0) \to H^p(K'_\beta)$, and since $\pi^*_{\beta 0}$ and $\pi^*_{\alpha 0}$ are isomorphisms onto, $\pi^*_{\beta\alpha} = \pi^*_{\beta 0}\pi^*_{\alpha 0}{}^{-1}: H^p(K'_\alpha) \to H^p(K'_\beta)$ is also an isomorphism onto.

Since the coverings $\mathfrak{U}_\alpha$ form a cofinal family of coverings of $|K|$, the Čech cohomology group $H^p(|K|, G)$ of $|K|$ based on a discrete coefficient group G is the limit group of a direct spectrum, the groups of which are the cohomology groups $H^p(K'_\alpha, G)$ of the simplicial subdivisions of K_0 and the homomorphisms of which are the homomorphisms $\pi^*_{\beta\alpha}: H^p(K'_\alpha, G) \to H^p(K'_\beta, G)$. Since by (2.3) each $H^p(K'_\alpha, G)$ is isomorphic with $H^p(K, G)$, and since each $\pi^*_{\beta\alpha}$ is an isomorphism onto, it follows that the limit group $H^p(|K|, G)$ is also isomorphic with $H^p(K, G)$. Thus the Čech cohomology group $H^p(|K|, G)$ of K is isomorphic with the combinatorial cohomology group $H^p(K, G)$.

Another consequence of (7.1) is the following result.

(8.2) *The singular homology* (*cohomology*) *groups of a Whitehead complex K are isomorphic with the corresponding combinatorial homology* (*cohomology*) *groups of K.*

For the proof in case K is simplicial see ([7], pp. 399-400). If K is not simplicial, we replace it by a simplicial subdivision K'_0 where, by (6.1), $|K|$ and $|K'_0|$ are homeomorphic and, by (2.3), the combinatorial homology and cohomology groups of K are isomorphic with those of K'_0.

IV. Metric complexes.

9. Definition and properties of metric complexes. A metric complex is a topological complex whose underlying space is a metric space (a topological complex whose underlying space is metrizable will be called a metrizable complex). Replacing condition (b) of section **3** by the equivalent condition (b′), we can say that a metric complex is an affine complex K whose underlying set is a metric space subject to the conditions

(a) Each $\phi_\alpha: \bar{E}_\alpha \to \bar{e}_\alpha$ is a homomorphism,

(b′) For each point x of K and each positive number ϵ, there exists a positive number δ such that[9] if $\rho(x,y) < \delta$, then $x \varepsilon \bar{e}(y)$, and for each point z of the segment $[x,y]$ in $\bar{e}(y)$, $\rho(x,z) < \epsilon$.

Every subcomplex of a metric complex is a metric complex, for every subspace of a metric space is a metric space, and every subcomplex of a topological complex is a topological complex. The cartesian product of two metric complexes is a metric complex, for the product space of two metric spaces is a metric space, and the product of two topological complexes is a topological complex.

Let K be a metric complex. We define $\eta(x,\epsilon)$ as follows: Let $2\eta(x,\epsilon)$ be the least upper bound of the δ's of condition (b′) if this least upper bound exists and is less than ϵ; if the δ's are unbounded, or if their least upper bound is not less than ϵ, let $2\eta(x,\epsilon) = \epsilon$. Thus $\eta(x,\epsilon)$ is defined for each $x \varepsilon K$ and each $\epsilon > 0$, and $0 < \eta(x,\epsilon) \leqq \epsilon/2$. Clearly if $\epsilon_1 < \epsilon_2$, then $\eta(x,\epsilon_1) \leqq \eta(x,\epsilon_2)$. If $y \varepsilon K$, and $\rho(x,y) < 2\eta(x,\epsilon)$, then the segment $[x,y]$ exists in $\bar{e}(y)$, and for each $z \varepsilon [x,y]$, $\rho(x,z) < \epsilon$.

We define $\eta_r(x,\epsilon)$ inductively as follows, for $r = 0, 1, 2, \cdots$. Let $\eta_0(x,\epsilon) = \epsilon$, and for $r \geqq 1$, let $\eta_r(x,\epsilon) = \eta(x, \eta_{r-1}(x,\epsilon))$. One sees immediately that $\eta_1(x,\epsilon) = \eta(x,\epsilon)$, and $\eta_2(x,\epsilon) = \eta(x,\eta(x,\epsilon))$. It also follows from the definition that for $r > 0$, $\eta_r(x,\epsilon) \leqq \frac{1}{2}\eta_{r-1}(x,\epsilon)$, and for $r \geqq 2$, $\eta_r(x,\epsilon) \leqq \frac{1}{2}\eta(x,\epsilon)$.

(9.1) *For any metric complex, if $r \geqq 1$, then $\eta_r(x,\epsilon) = \eta_{r-1}(x,\eta(x,\epsilon))$.*

Proof. This is clear for $r = 1$. We proceed by induction. Let $r \geqq 2$, and assume $\eta_{r-1}(x,\epsilon) = \eta_{r-2}(x,\eta(x,\epsilon))$. Then $\eta_r(x,\epsilon) = \eta(x,\eta_{r-1}(x,\epsilon)) = \eta(x,\eta_{r-2}(x,\eta(x,\epsilon))) = \eta_{r-1}(x,\eta(x,\epsilon))$.

We write $S(x,\epsilon)$, or $S_0(x,\epsilon)$, for the set of points y of K such that $\rho(x,y) < \epsilon$, and $S_r(x,\epsilon)$ for the neighborhood $S(x,\eta_r(x,\epsilon))$. We write $\{p\}$ for the set consisting of one element p.

(9.2) *Let A be a set of r points ($r \geqq 1$) of a closed cell $\bar{e}_a$ of a metric complex K, let ϵ be a positive number, and let $\bar{x}$ be a point of A such that $\eta(\bar{x},\epsilon) \geqq \eta(x,\epsilon)$ for all $x \varepsilon A$. Then if there is some point y of $\bar{e}_a$ such that $y \varepsilon S_r(x,\epsilon)$ for all $x \varepsilon A$, the convex hull of $\{y\} \cup A$ in $\bar{e}_a$ is contained in $S(\bar{x},\epsilon)$.*

Proof. First let $r = 1$. Then $A = \{\bar{x}\}$, and the convex hull of $\{y\} \cup A$ in $\bar{e}_a$ is the segment $[\bar{x},y]$ in $\bar{e}_a$. Since $y \varepsilon S_1(\bar{x},\epsilon) = S(\bar{x},\eta(\bar{x},\epsilon))$, $\rho(\bar{x},y) < \eta(\bar{x},\epsilon)$, and hence if $z \varepsilon [\bar{x},y]$, $\rho(\bar{x},z) < \epsilon$. Thus $[\bar{x},y] \subset S(\bar{x},\epsilon)$.

[9] We write $\rho(x,y)$ for the distance between the points x and y.

We proceed by induction. Let $r \geq 2$, and assume the result has been proved for $r-1$. Let $B = A - \{\bar{x}\}$, let $B_1 = \{y\} \cup B$, and let $A_1 = \{y\} \cup A$. Then for any point z of the convex hull $A_1{}^*$ of A_1 in $\bar{e}_\alpha$, $z \,\varepsilon\, [\bar{x}, y_1]$, where y_1 is a point in the convex hull $B_1{}^*$ of B_1 in $\bar{e}_\alpha$. By the induction hypothesis, since $y \,\varepsilon\, S_{r-1}(x, \eta(x, \epsilon)) \subset S_{r-1}(x, \eta(\bar{x}, \epsilon))$ for each $x \,\varepsilon\, B$, $B_1{}^* \subset S(\tilde{\bar{x}}, \eta(\bar{x}, \epsilon))$ for some $\tilde{\bar{x}} \,\varepsilon\, B$. Hence $\rho(\tilde{\bar{x}}, y_1) < \eta(\bar{x}, \epsilon)$, and $\rho(\bar{x}, y_1) \leq \rho(\bar{x}, y) + \rho(y, \tilde{\bar{x}}) + \rho(\tilde{\bar{x}}, y_1) < \eta_r(\bar{x}, \epsilon) + \eta_r(\tilde{\bar{x}}, \epsilon) + \eta(\bar{x}, \epsilon) < \frac{1}{2}\eta(\bar{x}, \epsilon) + \frac{1}{2}\eta(\tilde{\bar{x}}, \epsilon) + \eta(\bar{x}, \epsilon) \leq 2\eta(\bar{x}, \epsilon)$. Therefore $\rho(x, z) < \epsilon$, and $A_1{}^* \subset S(\bar{x}, \epsilon)$.

(9.3) *For $r \geq 1$ let A be a set of r points, not necessarily distinct, of a closed cell $\bar{e}_\alpha$ of a metric complex K, and let ϵ be a positive number. Then if C is a non-empty convex set of $\bar{e}_\alpha$ contained in $\bigcap_{x \varepsilon A} S_r(x, \epsilon)$, the convex hull of $C \cup A$ in $\bar{e}_\alpha$ has diameter less than 2ϵ.*

Proof. Note that $(C \cup A)^*$ is the union of the sets $(\{y\} \cup A)^*$ for all $y \,\varepsilon\, C$. By Lemma 2, each $(\{y\} \cup A)^* \subset S(\bar{x}, \epsilon)$, where $\bar{x}$ is a point of A for which $\eta(\bar{x}, \epsilon)$ is maximal. Hence $(C \cup A)^* \subset S(\bar{x}, \epsilon)$, and the diameter of $(C \cup A)^*$ is less than 2ϵ.

10. The conditions to be satisfied. Let K be a metric complex. For each positive integer n we choose a collection $\mathfrak{U}^n$ of open sets of K, and with each open set U_λ of $\mathfrak{U}^n$ we associate a point x_λ of K. (The indices λ of the sets of $\mathfrak{U}^n$ are assumed to be elements of some sufficiently large index set.) With each cell e_α of K and each positive integer n, we associate a subcollection $\mathfrak{U}_\alpha{}^n$ of $\mathfrak{U}^n$ and a positive real number $\rho_\alpha{}^n$. The choice of $\mathfrak{U}^n$, x_λ, $\mathfrak{U}_\alpha{}^n$ and $\rho_\alpha{}^n$ will be subject to the following conditions:

1) $\mathfrak{U}^n = \bigcup_\alpha \mathfrak{U}_\alpha{}^n$

2) $\mathfrak{U}_\alpha{}^n$ is a finite collection of open sets whose union contains $\bar{e}_\alpha$.

3) If $U_\lambda \,\varepsilon\, \mathfrak{U}_\alpha{}^n$ then $x_\lambda \,\varepsilon\, \bar{e}_\alpha$ and, for any $y \,\varepsilon\, U_\lambda$ and $z \,\varepsilon\, K - \operatorname{St} e(x_\lambda)$, $\rho(x_\lambda, y) < \frac{1}{2}\rho(x_\lambda, z)$.

4) If $e_\beta < e_\alpha$, $\mathfrak{U}_\beta{}^n \subset \mathfrak{U}_\alpha{}^n$.

5) If, for $e_\beta \leq e_\alpha$, $U_\lambda \,\varepsilon\, \mathfrak{U}_\alpha{}^n$, and $x_\lambda \,\varepsilon\, U_\mu \,\varepsilon\, \mathfrak{U}_\beta{}^n$, then $U_\lambda \,\varepsilon\, \mathfrak{U}_\beta{}^n$.

6) Let $e_\alpha \leq e_\gamma$, and let C be a convex set in $\bar{e}_\gamma$. Let $U_{\lambda_0}, \cdots, U_{\lambda_p}$ be sets of $\mathfrak{U}_\alpha{}^n$, and let $U_{\mu_0}, \cdots, U_{\mu_q}$ be sets of $\mathfrak{U}_\alpha{}^{n+1}$ such that the intersection

$$C \cap U_{\lambda_0} \cap \cdots \cap U_{\lambda_p} \cap U_{\mu_0} \cap \cdots \cap U_{\mu_q}$$

is not empty. Then if C has diameter less than $2\rho_\alpha{}^n$, the convex hull of the union $C \cup \{x_{\lambda_0}, \cdots, x_{\mu_q}\}$ in $\bar{e}_\gamma$ has diameter less than $1/n$.

Note that condition 3) implies condition 3′).

3′) If $U_\lambda \varepsilon \mathfrak{U}_\alpha{}^n$, then $e(x_\lambda) \leq e_\alpha$, and $U_\lambda \subset \operatorname{St} e(x_\lambda)$.

Also, if condition 3) holds for each of two cells e_α and e_β, then condition 3″) holds.

3″) If $U_\lambda \varepsilon \mathfrak{U}_\alpha{}^n$ and $U_\mu \varepsilon \mathfrak{U}_\beta{}^m$, and if $U_\lambda \cap U_\mu \neq 0$, then either

$$e(x_\mu) > e(x_\lambda) \quad \text{or} \quad e(x_\mu) \leq e(x_\lambda).$$

For if neither of $e(x_\lambda)$ and $e(x_\mu)$ is a face of the other, then $x_\lambda \varepsilon e(x_\lambda) \subset K - \operatorname{St} e(x_\mu)$, and $x_\mu \varepsilon e(x_\mu) \subset K - \operatorname{St} e(x_\lambda)$. Let $x \varepsilon U_\lambda \cap U_\mu$. By 3) applied to e_α, $\rho(x_\lambda, x) < \frac{1}{2}\rho(x_\lambda, x_\mu)$, and by 3) applied to e_β, $\rho(x_\mu, x) < \frac{1}{2}\rho(x_\mu, x_\lambda)$. Therefore $\rho(x_\lambda, x_\mu) > \rho(x_\lambda, x) + \rho(x_\mu, x)$, which is absurd.

Also, if conditions 2) and 5) hold for e_α and all of its faces, the following condition 5′) also holds.

5′) If, for $e_\beta \leq e_\alpha$, $U_\lambda \varepsilon \mathfrak{U}_\alpha{}^n$ and $x_\lambda \varepsilon \bar{e}_\beta$, then $U_\lambda \varepsilon \mathfrak{U}_\beta{}^n$.

For by 2). $\mathfrak{U}_\beta{}^n$ covers $\bar{e}_\beta$, and hence for some $U_\mu \varepsilon \mathfrak{U}_\beta{}^n$, $x_\lambda \varepsilon U_\mu$. Thus by 5), $U_\lambda \varepsilon \mathfrak{U}_\beta{}^n$.

11. The construction. We shall first construct the collections $\mathfrak{U}_\alpha{}^n$ of open sets, and afterwards we shall use condition 1 to define $\mathfrak{U}^n$. The construction of $\mathfrak{U}_\alpha{}^n$ will be by induction on the dimension of the cell e_α.

First let e_α be a cell of dimension zero; that is, e_α consists of a single point v. If v is the only point of K, let $\mathfrak{U}_\alpha{}^n$ consist of the one set $U_\lambda = \{v\}$, let x_λ be the point v, and let $\rho_\alpha{}^n = 1$. If K has other points, and hence other vertices, let $d(v)$ be the distance from v to the complement of the open set $\operatorname{St} e_\alpha$, and let $\mathfrak{U}_\alpha{}^n$ consist of a single open set U_λ which is the spherical neighborhood of v with radius the smaller of $\frac{1}{2}d(v)$ and $\eta(v, 1/2n)$. Let x_λ be the point v, and let $2\rho_\alpha{}^n = \eta(v, 1/2n)$. In either case the proof that conditions 2) to 6) are satisfied is easy, and is left to the reader.

Now suppose that e_α has dimension $k > 0$, and suppose that $\mathfrak{U}_\beta{}^n$, x_λ, and $\rho_\beta{}^n$ satisfying conditions 2) to 6) have been constructed for all cells e_β of dimenion less than k (in particular for all proper faces e_β of e_α), and for all n.

For each n, let $A_\alpha{}^n$ be the set of points of $\bar{e}_\alpha$ which, for no proper face e_β of e_α, are contained in a set of the covering $\mathfrak{U}_\beta{}^n$. Then $A_\alpha{}^n$ is a closed and hence compact subset of $\bar{e}_\alpha$, and $A_\alpha{}^n \subset e_\alpha$. Let $\theta_\alpha{}^n$ be the least value of $\rho_\beta{}^m$ for all $e_\beta < e_\alpha$ and all $m \leq n$. Since e_α has only a finite number of faces, and since a finite number of positive integers precede $n + 1$, the number $\theta_\alpha{}^n$ exists and is positive. For each point x of $A_\alpha{}^n$, let $d(x)$ be the distance from

x to the complement of St e_α, and let $N(x)$ be the spherical neighborhood of x with radius $r(x)$ equal to the smaller of $\frac{1}{2}d(x)$ and $(1/3)\eta_{2k+2}(x, \theta_\alpha{}^n)$. Since $A_\alpha{}^n$ is compact and of dimension $\leq k$, the covering of A^n by the neighborhoods $N(x)$ has as a refinement a finite covering $\mathfrak{B}_\alpha{}^n$ by open sets of K, such that no point of K is contained in more that $k+1$ sets of $\mathfrak{B}_\alpha{}^n$. For each set U_λ of $\mathfrak{B}_\alpha{}^n$, let x_λ be any one of the points of $A_\alpha{}^n$ for which $N(x_\lambda) \supset U_\lambda$. Let $\kappa_\alpha{}^n$ be the least of the finite number of distances $r(x)$, more explicitly $r^n(x)$, corresponding to the finite number of U_λ in $\mathfrak{B}_\alpha{}^n$; and let $\rho_\alpha{}^n$ be the smaller of the two positive numbers $\kappa_\alpha{}^n$ and $\kappa_\alpha{}^{n+1}$. Let $\mathfrak{U}_\alpha{}^n$ consist of the sets of $\mathfrak{B}_\alpha{}^n$ together with all sets of $\mathfrak{U}_\beta{}^n$ for all proper faces e_β of e_α.

12. Verification of the conditions. We now verify that the $\mathfrak{U}_\alpha{}^n$, x_λ, and $\rho_\alpha{}^n$ so defined satisfy conditions 2) to 6).

2) The collection $\mathfrak{U}_\alpha{}^n$ is the union of the finite collection $\mathfrak{B}_\alpha{}^n$ and the finite number of finite collections $\mathfrak{U}_\beta{}^n$ for $e_\beta < e_\alpha$. Hence $\mathfrak{U}_\alpha{}^n$ is finite. For any $x \in \bar{e}_\alpha$, either $x \in U_\lambda \in \mathfrak{U}_\beta{}^n$ for some $e_\beta < e_\alpha$, or $x \in A_\alpha{}^n$ and is contained in a set of the covering $\mathfrak{B}_\alpha{}^n$. Hence in either case, x is contained in a set of $\mathfrak{U}_\alpha{}^n$. Thus $\bar{e}_\alpha$ is contained in the union of the collection $\mathfrak{U}_\alpha{}^n$ of open sets, and condition 2) is satisfied.

3) If $U_\lambda \in \mathfrak{B}_\alpha{}^n$, then $x_\lambda \in A_\alpha{}^n \subset e_\alpha \subset \bar{e}_\alpha$. Then, if $y \in U_\lambda \subset N(x_\lambda)$, $\rho(x_\lambda, y) < r(x_\lambda) \leq \frac{1}{2}d(x_\lambda)$, and if $z \in K -$ St $e(x_\lambda) = K -$ St e_α, $\rho(x_\lambda, z) \geq d(x_\lambda)$; hence $\rho(x,y) < \frac{1}{2}\rho(x,z)$. If $U_\lambda \in \mathfrak{U}_\alpha{}^n$ but $\notin \mathfrak{B}_\alpha{}^n$, then for some $e_\beta < e_\alpha$, $U_\lambda \in \mathfrak{U}_\beta{}^n$. Applying condition 3) to the lower dimensional cell e_β, $x_\lambda \in \bar{e}_\beta \subset \bar{e}_\alpha$, and for $y \in U_\lambda$ and $z \in K -$ St $e(x_\lambda)$, $\rho(x_\lambda, y) < \frac{1}{2}\rho(x_\lambda, z)$. Thus condition 3) is satisfied.

4) By the definition of $\mathfrak{U}_\alpha{}^n$ condition 4) is satisfied.

5) Condition 5) is trivial for $e_\beta = e_\alpha$. Assume $e_\beta < e_\alpha$, $U_\lambda \in \mathfrak{U}_\alpha{}^n$, and $x_\lambda \in U_\mu \in \mathfrak{U}_\beta{}^n$. Then $x_\lambda \notin A_\alpha{}^n$, and hence $U_\lambda \notin \mathfrak{B}_\alpha{}^n$. Thus for some $e_\gamma < e_\alpha$, $U_\lambda \in \mathfrak{U}_\gamma{}^n$. Let $e_\delta = e(x_\mu)$. Then, by condition 3′) applied to e_β, since $U_\mu \in \mathfrak{U}_\beta{}^n$, we have $e_\delta \leq e_\beta$, and $U_\mu \subset$ St e_δ. By condition 5′) applied to e_δ and e_β, since $U_\mu \in \mathfrak{U}_\beta{}^n$, and $x_\mu \in \bar{e}_\delta$, we have $U_\mu \in \mathfrak{U}_\delta{}^n$. By 3′) applied to e_γ, since $U_\lambda \in \mathfrak{U}_\gamma{}^n$, we have $e(x_\lambda) \leq e_\gamma$. But $x_\lambda \in U_\mu \subset$ St e_δ; hence $e_\delta \leq e(x_\lambda)$, and $e_\delta \leq e_\gamma$. .By 5) applied to e_δ and e_γ, since $U_\lambda \in \mathfrak{U}_\gamma{}^n$ and $x_\lambda \in U_\mu \in \mathfrak{U}_\delta{}^n$, we have $U_\lambda \in \mathfrak{U}_\delta{}^n$. Since $e_\delta \leq e_\beta$, it follows from condition 4) that $\mathfrak{U}_\delta{}^n \subset \mathfrak{U}_\beta{}^n$, Hence $U_\lambda \in \mathfrak{U}_\beta{}^n$, and condition 5) is satisfied.

6) Let $e_\alpha \leq e_\gamma$, and let C be a convex set in $\bar{e}_\gamma$ with diameter $< 2\rho_\alpha{}^n$. Let $y \in C \cap U_{\lambda_0} \cap \cdots \cap U_{\lambda_p} \cap U_{\mu_0} \cap \cdots \cap U_{\mu_q}$, where each U_λ is in $\mathfrak{U}_\alpha{}^n$, and each U_μ is in $\mathfrak{U}_\alpha{}^{n+1}$. Let u_n be the set $(x_{\lambda_0}, \cdots, x_{\lambda_p})$, and let u_{n+1} be the set $(x_{\mu_0}, \cdots, x_{\mu_q})$.

Each point of K, and in particular y, is contained in at most $k+1$ sets of $\mathfrak{V}_\alpha{}^n$. Hence at most $k+1$ of the sets $U_{\lambda_0}, \cdots, U_{\lambda_p}$ are in $\mathfrak{V}_\alpha{}^n$, and at most $k+1$ of the points of u^n are in e_α. Similarly, at most $k+1$ of the points of u^{n+1} are in e_α. Let $v^n = u^n \cap e_\alpha$, $v^{n+1} = u^{n+1} \cap e_\alpha$; then $v^n \cup v^{n+1}$ has at most $2k+2$ points.

If $x_\lambda \varepsilon v^n$, then $y \varepsilon U_\lambda \subset N(x_\lambda)$, and hence $\rho(x_\lambda, y) < r^n(x_\lambda)$. Also diam $C < 2\rho_\alpha{}^n \leq 2\kappa_\alpha{}^n \leq 2r^n(x_\lambda)$. Hence, if $z \varepsilon C$,

$$\rho(x_\lambda, z) \leq \rho(x_\lambda, y) + \rho(y, z) < 3r^n(x_\lambda) \leq \eta_{2k+2}(x_\lambda, \theta_\alpha{}^n).$$

If $x_\mu \varepsilon v^{n+1}$, then $y \varepsilon U_\mu \subset N(x_\mu)$, and hence $\rho(x_\mu, y) < r^{n+1}(x_\mu)$. Also diam $C < 2\rho_\alpha{}^n \leq 2\kappa_\alpha{}^{n+1} \leq 2r^{n+1}(x_\mu)$. Hence if $z \varepsilon C$,

$$\rho(x_\mu, z) < 3r^{n+1}(x_\mu) \leq \eta_{2k-2}(x_\mu, \theta_\alpha{}^{n+1}).$$

By definition, $\theta_\alpha{}^n$ is a decreasing function of n; hence $\theta_\alpha{}^{n+1} \leq \theta_\alpha{}^n$, and $\eta_{2k+2}(x_\mu, \theta_\alpha{}^{n+1}) \leq \eta_{2k+2}(x_\mu, \theta_\alpha{}^n)$. Thus for each $z \varepsilon C$, $\rho(x_\mu, z) < \eta_{2k+2}(x_\mu, \theta_\alpha{}^n)$.

It follows that for each of the at most $2k+2$ points x of $v^n \cup v^{n+1}$, C is contained in the $\eta_{2k+2}(x, \theta_\alpha{}^n)$ neighborhood of x. Hence by (8.3), the diameter of the convex hull of $C \cup v^n \cup v^{n+1}$ in $\bar{e}_\gamma$ is less than $2\theta_\alpha{}^n$.

Let $w^n = u^n - v^n$, and let $w^{n+1} = u^{n+1} - v^{n+1}$. Then for each $x_\nu \varepsilon w^n \cup w^{n+1}$, $e(x_\nu) < e_\alpha$. Let x_π be chosen in $w^n \cup w^{n+1}$ so that $e(x_\pi)$ has maximum dimension, and let $e_\beta = e(x_\pi)$. Then if $x_\nu \varepsilon w^n \cup w^{n+1}$, $U_\pi \cap U_\nu \neq 0$, and hence by 3″) applied to the faces $e(x_\pi)$ and $e(x_\nu)$ of e_α, either $e(x_\nu) > e(x_\pi)$, or $e(x_\nu) \leq e(x_\pi)$. Since $e(x_\pi)$ is of maximal dimension, $e(x_\nu) > e(x_\pi)$ is impossible; hence $e(x_\nu) \leq e(x_\pi)$, and $x_\nu \varepsilon \bar{e}(x_\pi) = \bar{e}_\beta$. Therefore $w^n \cup w^{n+1} \subset \bar{e}_\beta$.

Let $C' = (C \cup v^n \cup v^{n+1})^*$, the convex hull of $C \cup v^n \cup v^{n+1}$. Then, diam $C' < 2\theta_\alpha{}^n \leq 2\rho_\beta{}^n$. Now $C' \subset \bar{e}_\gamma$, $e_\beta < e_\gamma$, and y is in C' and in each of the sets U_ν of $(U_{\lambda_0}, \cdots, U_{\mu_q})$ for which $x_\nu \varepsilon w^n \cup w^{n+1} \subset \bar{e}_\beta$. But by condition 5′), each set U_ν is in $\mathfrak{U}_\beta{}^n \cup \mathfrak{U}_\beta{}^{n+1}$. Thus we can apply condition 6) to e_β, and we find that $(C' \cup w^n \cup w^{n+1})^*$ has diameter less than $1/n$. But

$$\begin{aligned}(C' \cup w^n \cup w^{n+1})^* &= ((C \cup v^n \cup v^{n+1})^* \cup w^n \cup w^{n+1})^* \\ &= (C \cup v^n \cup v^{n+1} \cup w^n \cup w^{n+1})^* = (C \cup u^n \cup u^{n+1})^*.\end{aligned}$$

Therefore $(C \cup u^n \cup u^{n+1})^* = (C \cup \{x_{\lambda_0}, \cdots, x_{\mu_q}\})^*$ has diameter less than $1/n$, and condition 6) is satisfied.

Thus for each e_α of dimension k, we construct $\mathfrak{U}_\alpha{}^n$, $\{x_\lambda\}$, and $\rho_\alpha{}^n$ satisfying conditions 2) to 6). By induction on k, we have a family of $\mathfrak{U}_\alpha{}^n$, x_λ, and $\rho_\alpha{}^n$ satisfying conditions 2) to 6) for all e_α and all n. Let $\mathfrak{U}^n = \bigcup_\alpha \mathfrak{U}_\alpha{}^n$. Then all the conditions 1) to 6) are satisfied.

13. The families of open sets as coverings. We first verify that each family $\mathfrak{U}^n$ is a covering and in fact a locally finite covering of the metric complex K.

(13.1) *Each of the families* $\mathfrak{U}^n$, $n = 1, 2, \cdots$, *of open sets is a locally finite covering of* K.

Proof. Let $x \in K$, and let $e_a = e(x)$. Then by condition 2), x is contained in some set of $\mathfrak{U}_a{}^n \subset \mathfrak{U}^n$. Therefore $\mathfrak{U}^n$ is a covering of K. Let W_a be the union of the open sets of $\mathfrak{U}_a{}^n$; then W_a is an open set containing x. Let $a(x)$ be a positive number less than half the distance from x to the complement of the open set $W_a \cap St\, e_a$, and let $G(x)$ be the spherical neighborhood of x with radius $a(x)$.

Let $U_\lambda \in \mathfrak{U}^n - \mathfrak{U}_a{}^n$, and let $e_\gamma = e(x_\lambda)$. Then for some e_δ, $U_\lambda \in \mathfrak{U}_\delta{}^n$, and $x_\lambda \in \bar{e}_\delta$. Hence by 5′), $U_\lambda \in \mathfrak{U}_\gamma{}^n$. Since $U_\lambda \notin \mathfrak{U}_a{}^n$, e_γ is not a face of e_a, and so $e_a = e(x)$ is not in the star of $e_\gamma = e(x_\lambda)$. Hence $x \in K - \mathrm{St}\, e(x_\lambda)$.

Suppose it possible that $G(x) \cap U_\lambda \neq 0$, and let $y \in G(x) \cap U_\lambda$. Then by 3), since $y \in U_\lambda \in \mathfrak{U}_\gamma{}^n$, and $x \in K - \mathrm{St}\, e(x_\lambda)$, $\rho(x_\lambda, y) < \frac{1}{2}\rho(x_\lambda, x)$. Also, since $y \in G(x)$, $\rho(x, y) < a(x)$. Hence $\rho(x_\lambda, x) \leq \rho(x_\lambda, y) + \rho(x, y) < \frac{1}{2}\rho(x_\lambda, x) + a(x)$. Therefore $\rho(x_\lambda, x) < 2a(x)$, and hence $x_\lambda \in W_a \cap \mathrm{St}\, e_a$. Thus for some $U_\mu \in \mathfrak{U}_a{}^n$, $x_\lambda \in U_\mu$, and $e_a \leq e(x_\lambda) = e_\gamma$. Hence by 5), $U_\lambda \in \mathfrak{U}_a{}^n$, which is absurd. Therefore $G(x) \cap U_\lambda = 0$. Thus $G(x)$ meets only a finite number of open sets of $\mathfrak{U}^n$, $\mathfrak{U}^n$ is locally finite.

(13.2) *For any point* $x \in K$, *let* u *be the set of all* x_λ *for which* $x \in U_\lambda \in \mathfrak{U}^n$, *and let* u' *be the set of all* x_μ *for which* $x \in U_\mu \in \mathfrak{U}^{n+1}$. *Then* u *and* u' *are contained in* $\bar{e}(x)$, *and the convex hull of* $\{x\} \cup u \cup u'$ *in* $\bar{e}(x)$ *has diameter less than* $1/n$.

Proof. Let $e_a = e(x)$. and let C be the set $\{x\}$ consisting of one point. Then C is a convex set in $\bar{e}_a$, and $\operatorname{diam} C = 0 < 2\rho_a{}^n$. As in the proof of (13.1), there is a neighborhood $G(x)$ of x which meets no sets of $\mathfrak{U}^n - \mathfrak{U}_a{}^n$. Hence the sets $U_{\lambda_0}, \cdots, U_{\lambda_p}$ of $\mathfrak{U}^n$ which contain x are in $\mathfrak{U}_a{}^n$, and $u = \{x_{\lambda_0}, \cdots, x_{\lambda_p}\}$ is contained in $\bar{e}_a$. Similarly, the sets $U_{\mu_0}, \cdots, U_{\mu_q}$ of $\mathfrak{U}^{n+1}$ which contain x are in $\mathfrak{U}_a{}^{n+1}$, and hence $u' = \{x_{\mu_0}, \cdots, x_{\mu_q}\}$ is contained in $\bar{e}_a$. Then by condition 6), the convex hull of $C \cup u \cup u' = \{x\} \cup u \cup u'$ in $\bar{e}_a$ has diameter less than $1/n$.

14. The mappings of K into K. Corresponding to each of the coverings $\mathfrak{U}^n$ of K we define a mapping $\psi_n : N^n \to K$, where N^n is the Whitehead nerve

of $\mathfrak{U}^n$. For each vertex u_λ of the nerve N^n there is a corresponding open set $U_\lambda \varepsilon \mathfrak{U}^n$, and there is a point x_λ of K associated with U_λ; we define $\psi_n(u_\lambda) = x_\lambda$. For each simplex $\sigma = u_{\lambda_0} \cdots u_{\lambda_p}$ of N^n, there is some point $x \varepsilon K$ contained in the intersection $U_{\lambda_0} \cap \cdots \cap U_{\lambda_p}$. Among the cells of K which contain at least one of the points $x_{\lambda_0}, \cdots, x_{\lambda_p}$, let e_μ be one of maximum dimension. Then it follows from condition 3$''$) that e_μ is unique, $x_{\lambda_0}, \cdots, x_{\lambda_p}$ are all in $\bar{e}_\mu$, and by condition 3$'$), $e_\mu \leq e(x)$. We define $\psi_n | \sigma$ to be the linear map of σ into $\bar{e}_\mu$ determined by mapping each vertex u_{λ_i} of σ into the corresponding x_{λ_i}. Since $e_\mu \leq e(x)$, $\psi_n | \sigma$ is also a linear map of σ into $\bar{e}(x)$ for each $x \varepsilon U_{\lambda_0} \cap \cdots \cap U_{\lambda_p}$. If σ_1 is a face of σ, the point x is in each U_{λ_i} corresponding to a vertex u_{λ_i} of σ_1, and therefore $\psi_n | \sigma_1$ is also a linear map into $\bar{e}(x)$. Hence $\psi_n | \bar{\sigma}$ is a linear map into $\bar{e}(x)$, and $\psi_n | \bar{\sigma}$ is continuous. Since N^n has the Whitehead topology, $\psi_n : N^n \to K$ is therefore continuous.

Let K_W be the complex K retopologized with the Whitehead topology, and let $f : K_W \to K$ be the identity map; i. e. for each $x \varepsilon K_W$, let $f(x) = x \varepsilon K$. Since the topology of K_W is at least as fine as that of K, f is continuous. The map $\psi_n : N^n \to K$ can be factored into a map $\chi_n : N^n \to K_W$ followed by $f : K_W \to K$. Since χ_n is linear and hence continuous on each closed simplex $\bar{\sigma}$ of N^n, $\chi_n : N^n \to K_W$ is continuous. Let ϕ_n be a canonical map of K into the nerve N^n of $\mathfrak{U}^n$. Then ([3], p. 202) for each $x \varepsilon K$, $\phi_n(x) \varepsilon \bar{\sigma}(x)$, where $\sigma(x)$ is the simplex of N^n determined by x; the vertices of $\sigma(x)$ correspond to the open sets of $\mathfrak{U}^n$ containing x. If $\sigma(x) = u_{\lambda_0} \cdots u_{\lambda_p}$, then $x \varepsilon U_{\lambda_0} \cap \cdots \cap U_{\lambda_p}$, and hence ψ_n maps $\bar{\sigma}(x)$ into $\bar{e}(x)$. Hence $\psi_n \phi_n(x) \varepsilon \bar{e}(x)$. Let $g_n = \chi_n \phi_n :$ $K \to K_W$; then (see diagram) $f g_n = f \chi_n \phi_n = \psi_n \phi_n : K \to K$. Thus, for each $x \varepsilon K$, $f g_n(x) \varepsilon \bar{e}(x)$.

$$\begin{tikzcd} K \arrow[r, "\phi_n"] \arrow[d, "g_n"'] & N^n \arrow[d, "\psi_n"] \arrow[dl, dashed, "\chi_n"'] \\ K_W \arrow[r, "f"] & K \end{tikzcd}$$

15. The homotopy. We now define a homotopy $h : K \times I \to K$. For $n = 1, 2, \cdots$, let $h(x, 1/n) = f g_n(x) = \psi_n \phi_n(x)$. For $1/(n+1) < t < 1/n$, let $h(x, t)$ be the point of $\bar{e}(x)$ which divides the segment from $h(x, 1/(n+1))$ to $h(x, 1/n)$ in the ratio of $t - 1/(n+1)$ to $1/n - t$. Let $h(x, 0) = x$. We must show that $h : K \times I \to K$ is continuous.

Let $x \varepsilon K$, and let $e_a = e(x)$. Let V be an open set containing x which meets no sets of $\mathfrak{U}^n - \mathfrak{U}_a{}^n$ and meets no sets of $\mathfrak{U}^{n+1} - \mathfrak{U}_a{}^{n+1}$; then $V \subset \mathrm{St}\, e_a$. Let I_n be the interval $1/(n+1) \leq t \leq 1/n$. Let $(y, t) \varepsilon V \times I_n$. Then since $y \varepsilon V \subset \mathrm{St}\, e_a$, $e_a \leq e(y)$. If $y \varepsilon U_\lambda \varepsilon \mathfrak{U}^n$, we have $U_\lambda \varepsilon \mathfrak{U}_a{}^n$, and hence $x_\lambda \varepsilon \bar{e}_a$. Thus for each vertex u_λ of $\sigma(y)$ in N^n, $\psi_n(u_\lambda) \varepsilon \bar{e}_a$; hence ψ_n maps $\bar{\sigma}(y)$ into $\bar{e}_a \subset \bar{e}(y)$. Therefore, since $\phi_n(y) \varepsilon \bar{\sigma}(y)$, $fg_n(y) = \psi_n\phi_n(y) \varepsilon \bar{e}_a$. Similarly $fg_{n+1}(y) \varepsilon \bar{e}_a$. Since $e_a \leq e(y)$, the segment joining $fg_{n+1}(y)$ to $fg_n(y)$ in $\bar{e}(y)$ is the segment joining them in $\bar{e}_a$, and the point $h(y, t)$ is the point dividing the segment from $fg_{n+1}(y)$ to $fg_n(y)$ in $\bar{e}_a$ in the ratio $t - 1/(n+1)$ to $1/n - t$. Thus h maps $V \times I_n$ continuously into $\bar{e}_a \subset K$. It follows that for each n, h maps $K \times I_n$ continuously, and from this it follows that h is continuously except possibly for $t = 0$.

We must show that h is also continuous at points $(x, 0)$ of $K \times I$. For any $x \varepsilon K$ and $\epsilon > 0$, let U be the $(\epsilon/2)$-neighborhood of x in K, and let W be the subset $\{t \mid 0 \leq t < 1/m\}$ of I, where m is some integer greater than $2/\epsilon$. Then $U \times W$ is a neighborhood of $(x, 0)$ in $K \times I$. Let (y, t) be any point of $U \times W$. If $t = 0$, $h(y, t) = y$, $h(x, 0) = x$, and $\rho(h(x, 0), h(y, t)) = \rho(x, y) < \epsilon/2 < \epsilon$. If, on the other hand, $t > 0$, then for some $n \geq m$, $1/(n+1) \leq t \leq 1/n$. If $U_{\lambda_0}, \cdots, U_{\lambda_p}$ are the sets of $\mathfrak{U}^n$ containing y, then $\phi_n(y) \varepsilon \sigma(y)$, where $\sigma(y)$ is the simplex $u_{\lambda_0}, \cdots, u_{\lambda_p}$ of N^n. Since $\psi_n(u_{\lambda_i}) = x_{\lambda_i}$, and since $\psi_n \mid \bar{\sigma}(y)$ is a linear map into $\bar{e}(y)$, ψ_n maps $\bar{\sigma}(y)$ into the convex hull of $\{x_{\lambda_0}, \cdots, x_{\lambda_p}\}$ in $\bar{e}(y)$. Similarly if $U_{\mu_0}, \cdots, U_{\mu_q}$ are the sets of $\mathfrak{U}^{n+1}$ containing y, $\psi_{n+1}\phi_{n+1}(y) \varepsilon \{x_{\mu_0}, \cdots, x_{\mu_q}\}^*$ in $e(y)$. Therefore, since $h(x, t)$ is on the segment from $\psi_{n+1}\phi_{n+1}(y)$ to $\psi_n\phi_n(y)$ in $\bar{e}(y)$, $h(x, t) \varepsilon \{x_{\lambda_0}, \cdots, x_{\mu_q}\}^*$ in $\bar{e}(y)$. By (12. 2), diam $\{y, x_{\lambda_0}, \cdots, x_{\mu_q}\}^* < 1/n$; hence $\rho(y, h(x, t)) < 1/n \leq 1/m < \epsilon/2$. Because $y \varepsilon U$, $\rho(x, y) < \epsilon/2$. Hence, since $x = h(x, 0)$, $\rho(h(x, 0), h(y, t)) < \epsilon/2 + \epsilon/2 = \epsilon$. Thus h is continuous at $(x, 0)$, and therefore is continuous.

We have shown that $h : K \times I \to K$ is a homotopy.[10] Since $h(x, 0) = x$, and $h(x, 1) = fg_1(x)$, we have

(15. 1) *The map $fg_1 : K \to K$ is homotopic to the identity.*

We have in fact proved much more. We have shown that if $0 \leq t \leq 1/n$, then $\rho(y, h(y, t)) < 1/n$. Hence if $0 \leq t \leq 1$, $\rho(y, h(y, t/n)) < 1/n$. Thus if we set $h_n(x, t) = h(x, t/n)$, we have $h_n(x, 0) = h(x, 0) = x$, and $h_n(x, 1) = h(x, 1/n) = \psi_n\phi_n(x)$, and therefore h_n is a homotopy of the identity

[10] Actually the homotopy is a uniform homotopy. For definition of uniform homotopy see [3], p. 204.

map of K in K to the map $\psi_n\phi_n$ such that, for each $x \varepsilon K$ and $t \varepsilon I$, $\rho(x, h_n(x, t)) < 1/n$. Thus if we are given $\epsilon > 0$, and if we choose $n > 1/\epsilon$, then for every $t \varepsilon I$, $h_n(x, t)$ is within ϵ of x. We have therefore obtained

(15.2)[11] *Given any positive number ϵ, the identity map of the metric complex K on itself is ϵ-homotopic to a factored map $\psi_n\phi_n$, where ϕ_n is a map of K into a Whitehead complex N^n, and ψ_n is a map of N^n into K. During the homotopy a point x does not leave the closure $\bar{e}(x)$ of the cell containing it.*

16. The homotopy type. Two spaces X and Y are said to have the same *homotopy type* if there exist maps $f: X \to Y$ and $g: Y \to X$ such that $fg: Y \to Y$ is homotopic to the identity, and $gf: X \to X$ is homotopic to the identity. Spaces of the same homotopy type can not be distinguished by any of the invariants of algebraic topology. Our main theorem is that a metric complex and the corresponding Whitehead complex have the same homotopy type.

THEOREM 1. *If K is a metric*[12] *complex and K_W is the complex re-topologized with the Whitehead topology, then K and K_W have the same homotopy type.*

Proof. We have maps $f: K_W \to K$ and $g_1: K \to K_W$ such that, by (15.1), $fg_1: K \to K$ is homotopic to the identity. It is then sufficient to show that $g_1f: K_W \to K_W$ is homotopic to the identity. We define the homotopy $h: K_W \times I \to K_W$ as follows. For each $x \varepsilon K_W$, $g_1f(x) = g_1(x) \varepsilon \bar{e}(x)$; we define $h(x, t)$ to be the point which divides the segment from x to $g_1f(x)$ in $\bar{e}(x)$ in the ratio $t : 1 - t$. For each cell $\bar{e}_a$ of K_W, if $x \varepsilon \bar{e}_a$, then $e(x) \leq e_a$, and hence $h(x, t)$ is the point dividing the segment from x to $g_1f(x)$ in $\bar{e}_a$ in the ratio $t : (1 - t)$. Thus $h \mid \bar{e}_a \times I$ is continuous, and by (5.2), $h: K_W \times I \to K_W$ is a homotopy.

THEOREM 2. *Isomorphic metric complexes have the same homotopy type.*

Proof. Let K and L be isomorphic metric complexes. Then K_W and L_W are isomorphic Whitehead complexes. Hence by (6.2), K_W and L_W are homeomorphic and a fortiori of the same homotopy type. By Theorem 1,

[11] In the terminology of Lefschetz ([9], p. 98), (15.2) says that the identity mapping of K on itself is ϵ-deformable, for all ϵ, to a mapping into a continuous complex. We may interpret this as meaning that K is an absolute neighborhood retract.

[12] In this and the following theorems it is sufficient to assume that K is a metrizable complex.

K has the same homotopy type as K_W, and L has the same homotopy type as L_W. Hence K has the same homotopy type as L.

17. Canonical mappings. We now show that the theorems on canonical mappings [13] into the nerve of a covering hold equally whether the nerve is provided with a metric or with the Whitehead topology.

THEOREM 3. *Let X be a topological space, let $\mathfrak{U}$ be a covering of X, let N be the nerve of $\mathfrak{U}$ with the Whitehead topology, and let M be the nerve metrized in some way to form a metric complex. Then there exists a canonical map of X into the nerve N of $\mathfrak{U}$ if and only if there is a canonical map of X into the nerve M of $\mathfrak{U}$.*

Proof. First let ϕ be a canonical map of X into N, and let f be the identity map of N $(=M_W)$ onto M. Then $f\phi: X \to M$ is continuous. Also, for each vertex u_λ of M, $(f\phi)^{-1}\mathrm{St}_M(u_\lambda) = \phi^{-1}\mathrm{St}_N(u_\lambda) \subset U_\lambda$. Hence $f\phi$ is a canonical map of X into M.

On the other hand, let ψ be a canonical map of X into M. Let g be a map of M into N $(=M_W)$ (see § 14) such that, for each $x \varepsilon M$, $g(x) \varepsilon \bar{e}(x)$ in N. Then $g\psi$ is a map of X into N. If $g(x) \varepsilon e_\alpha$, $x \varepsilon \mathrm{St}\, e_\alpha$; hence $g^{-1}e_\alpha \subset \mathrm{St}\, e_\alpha$. If $e_\gamma \subset \mathrm{St}\, e_\alpha$, $g^{-1}e_\gamma \subset \mathrm{St}\, e_\gamma \subset \mathrm{St}\, e_\alpha$. Hence $g^{-1}\mathrm{St}_N e_\alpha \subset \mathrm{St}_M e_\alpha$. It follows that $(g\psi)^{-1}\,\mathrm{St}_N\, u_\lambda = \psi^{-1}g^{-1}\,\mathrm{St}_N\, u_\lambda \subset \psi^{-1}\,\mathrm{St}_M\, u_\lambda \subset U_\lambda$. Therefore $g\psi$ is a canonical map of X into N.

COROLLARY 1. *Let X be a normal space, and let $\mathfrak{U}$ be a covering of X. Then there is a canonical map of X into a metric nerve M of $\mathfrak{U}$, or into the Whitehead nerve N of $\mathfrak{U}$, if and only if $\mathfrak{U}$ has a locally finite refinement.*

COROLLARY 2. *Let X be a topological space. There is a canonical map of X into a metric nerve M or into the Whitehead nerve N of every covering $\mathfrak{U}$ of X if and only if X is paracompact and normal.*

Proof. These results have been proved for particular metric nerves ([4], p. 388). It follows from Theorem 3 that they hold for Whitehead nerves. By another application of Theorem 3 they hold for any other metric nerves.

18. Topological invariance. Finally we show the topological invariance invariance of the homology and cohomology groups of metric complexes.

THEOREM 4. *The combinatorial homology and cohomology groups of a*

[13] For definition of canonical mappings see section 7 above.

metric complex are isomorphic with the corresponding singular homology groups. If the coefficient group is discrete, the combinatorial cohomology groups are isomorphic with the Čech cohomology groups.

Proof. The singular homology and cohomology groups and the Čech cohomology groups are invariants of the homotopy type ([7], p. 400; [5], p. 287). Hence by Theorem 1 they are the same for K and for K_W. By (8.1) and (8.2) they are therefore isomorphic with the corresponding combinatorial homology and cohomology groups.

COROLLARY. *If K and L are homeomorphic metric complexes, their combinatorial homology and cohomology groups are isomorphic.*

HARVARD UNIVERSITY.

BIBLIOGRAPHY

[1] P. Alexandroff and H. Hopf, *Topologie*, Berlin (1935).

[2] N. Bourbaki, *Éléments de Mathematique* II, (Actualités scientifiques et industrielles, No. 858), Paris (1940).

[3] C. H. Dowker, "Mapping theorems for non-compact spaces," *American Journal of Mathematics*, vol. 69 (1947), pp. 200-242.

[4] ———, "An extension of Alexandroff's mapping theorem," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 386-391.

[5] ———, "Čech cohomology theory and the axioms," *Annals of Mathematics*, vol. 51 (1950), pp. 276-292.

[6] S. Eilenberg, "Singular homology theory," *Annals of Mathematics*, vol. 45 (1944), pp. 407-447.

[7] W. Hurewicz, J. Dugundji and C. H. Dowker, "Continuous connectivity groups in terms of limit groups," *Annals of Mathematics*, vol. 49 (1948), pp. 391-406.

[8] S. Lefschetz, *Algebraic Topology*, New York (1942).

[9] ———, *Topics in Topology*, Princeton (1942).

[10] K. Reidemeister, *Topologie der Polyeder*, Leipzig (1938).

[11] J. H. C. Whitehead, "Simplicial spaces, nuclei and m-groups," *Proceedings of the London Mathematical Society* (2), vol. 45 (1939), pp. 243-327.

[12] ———, "Combinatorial homotopy I," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 213-245.

ON THE UNBOUNDEDNESS OF THE ESSENTIAL SPECTRUM.*

By C. R. Putnam.

1. In the differential equation

$$x'' + (\lambda - f)x = 0, \tag{1}$$

let λ denote a real parameter, and let $f = f(t)$ be a real-valued continuous function on the half-line $0 \leqq t < \infty$. (Throughout this paper only real-valued functions will be considered.) In addition, suppose that f is such that the differential equation (1) is of the Grenzpunkt type, so that (1) possesses for some λ (and hence for every λ) a solution x which fails to belong to the class $L^2 = L^2[0, \infty)$; *cf.* Weyl [11], p. 238. In this case, the equation (1) and a linear, homogeneous boundary condition

$$x(0)\cos\alpha + x'(0)\sin\alpha = 0, \qquad 0 \leqq \alpha < \pi, \tag{2}$$

determine a boundary value problem on $0 \leqq t < \infty$ with a spectrum S_α. If S' denotes the (closed, possibly empty) set of cluster points of S_α, then S' is independent of α ([11], p. 251), and is called the essential spectrum of (1).

Various results concerning the set S' are known in case f is subject to certain restrictions. If, for instance, f satisfies

$$f(t) \to c \quad \text{as } t \to \infty, \tag{3}$$

then S' is the half-line $c \leqq t < \infty$; [3]. In general, the complement of S' is an open (possibly empty) set, and hence possesses a decomposition $\Sigma(\lambda_k, \lambda^k)$ into open intervals $\lambda_k < \lambda < \lambda^k$ (or "gaps" in S'), where it is understood that one, or possibly two, of the intervals are half-lines, and that the summation may consist of the single interval $-\infty < \lambda < \infty$ in case S' is empty. If f is bounded, so that

$$|f(t)| < \text{const.}, \qquad 0 \leqq t < \infty, \tag{4}$$

it is known that, except for $(\lambda_0, \lambda^0) = (-\infty, \lambda^0)$, the inequality $\lambda^k - \lambda_k \leqq$ const. holds, so that, in particular, S' is unbounded from above ([8]), and if certain extra conditions are placed on f, even asymptotic estimates of the gaps in S' can be given; *cf.* [6].

---

* Received September 15, 1951.

In most instances, the set S' is, if it is not empty, unbounded either from above or from below. However, examples of functions f for which S' consists of a single point are known; [4], Corollary 3, p. 110. Thus the set S' can be non-empty and yet bounded.

It is natural to ask under what general condition on f is the set S' unbounded when it is not empty. It turns out that a sufficient condition is that f be bounded from below, thus, f should satisfy

$$f(t) > \text{const.}, \qquad 0 \leqq t < \infty. \tag{5}$$

(It should be noticed that (5), and hence (4) or (3), imply that (1) is of the Grenzpunkt type; [11], p. 238.) Furthermore, as will be shown in Theorem I below, the restriction (5) even permits an asymptotic estimate of the gaps in S'. For convenience in the sequel, the following terminology will be introduced: Let $m_\alpha(\lambda) = \min |\lambda - \mu|$ when μ is in S_α, and let $m(\lambda) = \min |\lambda - \mu|$ when μ is in S'. Thus $m_\alpha(\lambda)$ and $m(\lambda)$ denote the distance from a fixed value λ to the nearest point of the set S_α or S'. Clearly $m(\lambda) = \infty$ for some λ (and hence for every λ) if and only if S' is empty. The following will be proved:

THEOREM I. *Let $f(t)$ be a real-valued continuous function on $0 \leqq t < \infty$ satisfying condition* (5). *Then one of the following two possibilities must occur: either* (i) *the set S' is empty or* (ii) *S' is not empty and is unbounded from above. Furthermore, in case* (ii), *$m(\lambda)$ satisfies*

$$m(\lambda) = O(\lambda^{\frac{1}{2}}), \qquad \lambda \to \infty. \tag{6}$$

The main tool used in this paper will be the lemma (*) of section **2** from which certain estimates for $m(\lambda)$ will be derived; various consequences of (*), in addition to Theorem I, will be set forth in Theorems II-IV in section **4**.

2. Consider the boundary value problem determined by (1) and (2) for a fixed α, and let $\lambda_1, \lambda_2, \cdots$ denote the eigenvalues (if any) and $\phi_1, \phi_2, \cdots$ the corresponding normalized eigenfunctions. Then for $\lambda = \lambda_j$ and $x = \phi_j$, equation (1) becomes

$$\phi_j'' + (\lambda_j - f)\phi_j = 0. \tag{7}$$

Let $L(x)$ be defined by $L(x) = x'' - fx$, and let g denote any function of class L^2 satisfying the boundary condition (2), for which $L(g)$ is defined, continuous, and of class L^2; thus,

$$\int_0^\infty g^2 dt < \infty, \qquad \int_0^\infty (L(g))^2 dt < \infty. \tag{8}$$

4

Since g satisfies (2), $\phi_j'(0)g(0) - \phi_j(0)g'(0) = 0$; in addition, by (8), $(\phi_j'(t)g(t) - \phi_j(t)g'(t)) \rightarrow 0$ as $t \rightarrow \infty$ (cf. [11], pp. 241-242). Multiplication of (7) by g followed by an integration readily leads to

$$(9) \qquad \int_0^\infty (L(g) + \lambda g)\phi_j dt = (\lambda - \lambda_j) \int_0^\infty g\phi_j dt,$$

for an arbitrary real number λ. Two applications of the Parseval relation applied to the functions $L(g) + \lambda g$ and g yield, in virtue of (9) and a similar relation in which the ϕ_j are replaced by the eigendifferentials corresponding to the continuous spectrum, the inequality

$$(10) \qquad \int_0^\infty (L(g) + \lambda g)^2 dt \geqq m_\alpha^2(\lambda) \int_0^\infty g^2 dt;$$

cf. [9], p. 140, for calculations of a similar nature.

First suppose that S' is not empty, so that $m(\lambda) < \infty$ (for all λ), and suppose $m_\alpha(\lambda) < m(\lambda)$. Let $g = g_n$, $n = 1, 2, \cdots$, denote any sequence of functions of the type considered above in the derivation of (10). In addition, suppose that

$$(11) \qquad \int_0^T g_n^2 dt / \int_0^\infty g_n^2 dt \rightarrow 0, \qquad n \rightarrow \infty,$$

holds for every fixed T satisfying $0 \leqq T < \infty$. Let ϵ denote a small positive number, and consider the λ-interval $[\lambda - m(\lambda) + \epsilon, \lambda + m(\lambda) - \epsilon]$. On this interval there exist at most a finite number of points $\lambda_1, \lambda_2, \cdots, \lambda_N$ (eigenvalues) in the spectrum of S_α. It follows from the Schwarz inequality and (11) that

$$(12) \qquad \sum_{j=1}^N \left(\int_0^\infty g_n\phi_j dt\right)^2 / \int_0^\infty g_n^2 dt \rightarrow 0, \qquad n \rightarrow \infty.$$

It is now an easy consequence of (12) that

$$\limsup_{n\rightarrow\infty} \int_0^\infty (L(g_n) + \lambda g_n^2) dt \geqq \limsup_{n\rightarrow\infty} (m(\lambda) - \epsilon)^2 \int_0^\infty g_n^2 dt,$$

whenever $\epsilon < m(\lambda)$ and the functions $g = g_n$ satisfy (11). Since ϵ is arbitrary, one can obtain even

$$(13) \qquad \limsup_{n\rightarrow\infty} \int_0^\infty (L(g_n) + \lambda g_n)^2 dt \geqq \limsup_{n\rightarrow\infty} [m^2(\lambda) \int_0^\infty g_n^2 dt].$$

Next, let y denote any function satisfying (2) for which y, y', and $L(y)$ are continuous and belong to L^2. Furthermore, let $\lambda > 0$, and put $h = \cos(\lambda^{\frac{1}{2}}t)$ and $g = yh$. It is readily verified that $g \cos\alpha + g' \sin\alpha = h(y \cos\alpha + y' \sin\alpha)$

$+ yh' \sin\alpha$, so that, since y satisfies (2) and since $h' = -\lambda^{\frac{1}{2}} \sin(\lambda^{\frac{1}{2}}t)$, g clearly satisfies (2). Furthermore, $L(g) + \lambda g = L(y)h + 2y'h'$, so that the left side of (10) becomes

$$\int_0^\infty (L(y)h + 2y'h')^2 dt. \tag{14}$$

One readily verifies as a consequence of the equation

$$\cos^2(\lambda^{\frac{1}{2}}t) = \tfrac{1}{2}(1 + \cos(2\lambda^{\frac{1}{2}}t))$$

and an integration by parts, that the integral on the right side of (10) is

$$\int_0^\infty y^2h^2 dt = \tfrac{1}{2}\int_0^\infty y^2 dt - \tfrac{1}{2}\lambda^{-\frac{1}{2}} \int_0^\infty yy' \sin(2\lambda^{\frac{1}{2}}t)\, dt. \tag{15}$$

(Use is made of the limit relation $y(t) \to 0$ as $t \to \infty$; this, in turn, is a consequence of the fact that y and y' belong to L^2.) An application of the Schwarz inequality to the second integral on the right side of equation (15) shows that

$$\int_0^\infty yy' \sin(2\lambda^{\frac{1}{2}}t)\, dt \leqq \left(\int_0^\infty y^2 dt \int_0^\infty y'^2 dt\right)^{\frac{1}{2}}. \tag{16}$$

If use is made of the inequality $(a+b)^2 \leqq 2(a^2 + b^2)$, where a and b are real, it is seen that (14) is majorized by

$$2\int_0^\infty [(L(y))^2 + 4\lambda y'^2]\, dt. \tag{17}$$

Furthermore, if y is normalized by

$$\int_0^\infty y^2 dt = 1, \tag{18}$$

it is seen from (15) and (16) that

$$\int_0^\infty y^2h^2 dt \geqq \tfrac{1}{2} - \tfrac{1}{2}\lambda^{-\frac{1}{2}}\left(\int_0^\infty y'^2 dt\right)^{\frac{1}{2}}. \tag{19}$$

Suppose now that $y = y_n$, $n = 1, 2, \cdots$, denotes any sequence of functions, satisfying the conditions imposed on y above, and such that (11), in which $g_n = y_n h$, is valid for every fixed T, where $0 \leqq T < \infty$. It then follows from (13) and the results obtained above that

$$4 \limsup_{n\to\infty} \int_0^\infty [(L(y_n))^2 + 4\lambda y_n'^2]\, dt$$
$$\geqq m^2(\lambda)[1 - \lambda^{-\frac{1}{2}} \liminf_{n\to\infty} \left(\int_0^\infty y_n'^2 dt\right)^{\frac{1}{2}}]. \tag{20}$$

It is clear from (19) that the condition (11) surely holds if

$$(21) \qquad 0 < \int_0^T y_n^2 dt / [1 - \lambda^{-\frac{1}{2}} (\int_0^\infty y_n'^2 dt)^{\frac{1}{2}}] \to 0, \qquad n \to \infty.$$

Furthermore, it is clear from the above discussion, that (20) is also valid in case S' is empty (so that $m(\lambda) = \infty$), with the understanding that the left side of the inequality of (20) may be ∞. The results obtained thus far can be summarized in the following lemma:

(*) *Let $\{y_n\}$ denote any sequence of real-valued functions on $0 \leqq t < \infty$, satisfying the boundary condition* (2) *for $x = y_n$ and the normalization condition* (18) *for $y = y_n$, and which, in addition, are such that y_n, y'_n and $L(y_n)$ are continuous, belong to L^2, and the inequality and limit relation of* (21) *are satisfied. Then the inequality* (20) *is valid for all $\lambda > 0$ (where $m(\lambda) \leqq \infty$).*

The above lemma will be used in the next section to prove Theorem I.

3. *Proof of Theorem* I. It is sufficient to show in this case that if S' is not empty, then (6) must hold. For it is an obvious consequence of (6) that S' must contain an infinity of points clustering at $\lambda = +\infty$ (and possibly elsewhere). But if S' is not empty, then S' is translated by c if $f(t)$ is replaced by $f(t) + c$, for any constant c. Since (5) remains valid for $f(t) + c$ whenever it holds for $f(t)$, it can be supposed that $\lambda = 0$ belongs to the set S'. Consider the fixed boundary value problem determined by (1) and (2) for $\alpha = 0$. Suppose first that $\lambda = 0$ is a cluster point of eigenvalues λ_n with corresponding normalized eigenfunctions ϕ_n, so that

$$(22) \qquad \int_0^\infty \phi_n^2 dt = 1.$$

It is clear that

$$(23) \qquad \int_0^\infty (L(\phi_n))^2 dt \to 0, \qquad n \to \infty.$$

Furthermire, multiplication of the equation (7) by ϕ_n for $j = n$, followed by an integration, yields

$$(24) \qquad \int_0^T \phi_n'^2 dt = \phi_n \phi_n' \Big|_0^T + \int_0^T (\lambda_n - f) \phi_n^2 dt,$$

for every T satisfying $0 \leqq T < \infty$. However, it follows from (5) that $-f <$ const., so that, since $\phi_n(0) = 0$,

$$\int_0^T \phi_n'^2 dt \leqq \phi_n(T) \phi_n'(T) + \text{const.}$$

It is known that $\phi_n(T)\phi_n'(T) \to 0$ as $T \to \infty$ (see the proof of the theorem of [12], p. 6; cf. also [2]) so that

$$(25) \qquad \int_0^\infty \phi_n'^2 dt \leqq \text{const.} \qquad \text{(independent of } n\text{)}.$$

Furthermore, the limit relation $\phi_n(t) \to 0$ as $n \to \infty$ holds uniformly on every finite t-interval $0 \leqq t \leqq T$; cf. [10], p. 269.[1] Hence, for sufficiently large λ, it follows from (25), that (21) holds for $y_n = \phi_n$. Moreover, it can be supposed that (for sufficiently large λ),

$$1 - \lambda^{-\frac{1}{2}} \limsup_{n\to\infty} \left(\int_0^\infty y_n'^2 dt \right)^{\frac{1}{2}} \geqq \tfrac{1}{2}$$

The lemma (*) is clearly applicable and shows that (20) holds and hence, by (22)-(24),

$$(26) \qquad \text{const.}\, \lambda \geqq \tfrac{1}{2} m^2(\lambda)$$

for large λ. That is, (6) is satisfied and the proof of Theorem I is complete, at least if $\lambda = 0$ is a cluster point of eigenvalues of the boundary value problem corresponding to $\alpha = 0$. If this last condition fails to hold, then $\lambda = 0$ is in the continuous spectrum of this boundary problem. (This alone would imply that S' is unbounded; [13].) It will be shown that again (6) must hold. (The proof is essentially similar to that carried out in case $\lambda = 0$ is a cluster point of eigenvalues.) One can obtain a sequence of functions y_n of the type considered in (*), so that, in particular, (25) holds if ϕ_n is replaced by y_n, and such that

$$(27) \qquad y_n'' + (\lambda_n - f) y_n = k_n,$$

where the k_n are continuous functions satisfying

$$(28) \qquad \int_0^\infty k_n^2 dt \to 0, \qquad n \to \infty;$$

see the lemma of [10], p. 269. In addition (23) is valid, and the limit relation $y_n(t) \to 0$, as $n \to \infty$ holds uniformly on every fixed t-interval $0 \leqq t \leqq T$; *loc. cit.*, p. 269. Finally, a relation of the type (25), but where ϕ_n is replaced by y_n, follows from (5), (27) and (28). Hence (26) can again be obtained and the proof of Theorem I is complete.

---

[1] Professor Wintner has pointed out to me that the passage occurring in [10], p. 267, referring to a classical theorem in [1], p. 278, should state that the methods of [1] imply that the class of an operator, on the L^2 space $0 \leqq t < \infty$, is unchanged by adding to it a bounded operator. Cf. a corresponding remark in [7], § 7.

4. In this section, a number of additional consequences of the lemma (*) of section **2** will be derived. First, a slight generalization of Theorem I is contained in the following

THEOREM II. *The essential spectrum S' is not empty and, in fact, holds when the assumption that f satisfies* (5) *is replaced by the assumption that* (1) *be of the Grenzpunkt type and that there exist a sequence of real-valued functions y_n satisfying the following three conditions*:

(i) *the y_n possess continuous second derivatives and satisfy the boundary condition* (2) *for a fixed α;*

(ii) $\int_0^\infty y_n^2 dt = 1$, $\int_0^T y_n^2 dt \to 0$ (*as $n \to \infty$, for every fixed positive number T*);

(iii) $\int_0^\infty y_n'^2 dt < \text{const.}$, $\int_0^\infty (L(y_n))^2 dt < \text{const.}$

That Theorem II implies Theorem I is clear. In fact, the conditions specified in II were obtained, in the proof of I, as a consequence of the assumption that (5) holds and that S' is not empty. The proof of III proceeds along the same lines as that of I, and can therefore be omitted.

THEOREM III. *If, in Theorem* II, *the first inequality in condition* (iii) *is strengthened to*

$$\int_0^\infty y_n'^2 dt \to 0, \qquad n \to \infty, \tag{29}$$

then the assertion (6) *of Theorem* II *can be improved to the statement*

$$m(\lambda) = O(1), \quad as\ \lambda \to \infty. \tag{30}$$

If, in addition to (29), *the second inequality of condition* (iii) *is replaced by*

$$\int_0^\infty (L(y_n))^2 dt \to 0, \qquad n \to \infty, \tag{31}$$

then the assertion of Theorem II *can be improved* to

(32) *the half-line $0 \leqq \lambda < \infty$ is contained in S'.*

In order to prove Theorem III one need note only that (20) now yields

$$\text{const.} \geqq m^2(\lambda) \qquad \text{or} \quad 0 \geqq m^2(\lambda), \tag{33}$$

according as (29) alone or both (29) and (31) are assumed. Since (33) is equivalent to (30) or (32), the proof of Theorem III is complete.

THEOREM IV. *Let $f(t)$ be a real-valued continuous function on $0 \leqq t < \infty$ satisfying* (5), *and suppose that, as $t \to \infty$, the (finite) value $\mu = \liminf f(t)$ belongs to the set S'. Then the set S' is precisely the half-line $\mu \leqq \lambda < \infty$.*

That S' is contained in the half-line $\mu \leqq \lambda < \infty$ is a consequence of the fact that the least point of S' is never less than μ; cf. [5], p. 850. It remains to show then that every $\lambda > \mu$ (hence every $\lambda \geqq \mu$) is in S'. To this end, as in the proof of Theorem I, it can be supposed that $f(t)$ is shifted, if necessary, by a constant, so that $\mu = 0$. Let $\lambda > 0$. It will be shown that there exists a sequence of functions y_n satisfying the conditions of Theorem III sufficient for the implication (32). For convenience it will be supposed that $\alpha = 0$ in (2) and that $\mu = 0$ is a cluster point of eigenvalues. (Note that the set S' is independent of α; moreover, the case in which $\mu = 0$ is not a cluster point of eigenvalues, and hence is in the continuous spectrum, can be treated as in the proof of Theorem I.) Again, one can obtain relation (23). It is clear from (24), the fact that $\mu = 0$, and from the properties of the functions ϕ_n occurring in the proof of Theorem I that

$$\limsup \int_0^\infty \phi_n'^2 dt = \limsup \left(-\int_0^\infty f\phi_n^2 dt\right) \leqq 0, \qquad (n \to \infty),$$

as an improvement to (25). The last formula line, for $y_n = \phi_n$, of course implies (29). It is now clear that the functions $y_n = \phi_n$ satisfy the conditions of Theorem III guaranteeing the validity of (32). This completes the proof of Theorem IV.

PURDUE UNIVERSITY.

REFERENCES.

[1] T. Carleman, *Sur les équations intégrales singulières à noyau réel et symétrique*, Uppsala, 1923.

[2] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[3] ———, "On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[4] ———, "Some examples in the theory of singular boundary value problems," *ibid., vol.* 74 (1952), pp. 107-126.

[5] ——— and C. R. Putnam, "The least cluster point of the spectrum of boundary value problems," *ibid.*, vol. 70 (1948), pp. 849-855.

[6] ——— and C. R. Putnam, "The gaps in the essential spectra of wave equations," *ibid.*, vol. 72 (1950), pp. 849-862.

[7] ——— and A. Wintner, "On perturbations of the continuous spectrum of the harmonic oscillator," *ibid.*, vol. 74 (1952), pp. 79-85.

[8] C. R. Putnam, "The cluster spectra of bounded potentials," *ibid.*, vol. 70 (1949), pp. 842-848.

[9] ———, "On isolated eigenfunctions associated with bounded potentials," *ibid.*, vol. 72 (1950), pp. 135-147.

[10] ———, "The comparison of spectra belonging to potentials with a bounded difference," *Duke Mathematical Journal*, vol. 18 (1951), pp. 267-273.

[11] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[12] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 72 (1947), pp. 5-13.

[13] ———, "On Dirac's theory of continuous spectra," *The Physical Review*, vol. 73 (1948), pp. 781-785.

PROPERTIES OF CONFORMAL INVARIANTS.*

By Vidar Wolontis.**

I. Basic Properties of Extremal Distance.

1 **Definition of extremal distance.**[1] Let D be a region in the complex plane, E_1 and E_2 two disjoint compact subsets of D, and Γ the class of all rectifiable curves in D joining E_1 and E_2. Let P be the set of non-negative functions $\rho(z)$ on D such that the integral

$$L_\rho(\gamma) = \int_\gamma \rho \mid dz \mid \tag{1}$$

is defined for all $\rho \varepsilon \mathrm{P}$ and every rectifiable curve γ in D, and

$$A_\rho(D) = \int\int_D \rho^2 dxdy, \qquad z = x + iy, \tag{2}$$

exists and is different from zero. We wish to determine $\rho \varepsilon \mathrm{P}$ so that the ratio

$$[\inf_{\gamma \varepsilon \Gamma} L_\rho(\gamma)]^2 / A_\rho(D) \tag{3}$$

be maximal. Since the existence of a maximum is not assured, we consider more generally the finite or positively infinite quantity

$$\lambda_D(E_1, E_2) = \sup_{\rho \varepsilon \mathrm{P}} [\inf_{\gamma \varepsilon \Gamma} L_\rho(\gamma)]^2 / A_\rho(D), \tag{4}$$

which we call the *extremal distance* between E_1 and E_2 with respect to the region D. For later use we observe that, since the value of the quotient (3) remains unchanged if ρ is multiplied by a positive constant, and since the extremal distance (4) is clearly never zero, we may restrict the class P e. g. by

* Received July 10, 1951.

** This paper includes the results found in my thesis, "Properties of Conformal Invariants," Harvard University, 1949. I wish to express my deep gratitude to Professor Lars V. Ahlfors for suggesting problems and methods, and for encouraging guidance and great personal interest.

[1] For the material presented in sections 1-3 of this chapter I am indebted to Professors Ahlfors and Beurling for their permission to consult an unpublished manuscript. Compare also A. Beurling, *Etudes sur un problème de majoration*, Thèse, Uppsala, 1933, and L. Ahlfors and A. Beurling, "Conformal invariants and function-theoretic null-sets," *Acta Mathematica*, vol. 83 (1950), pp. 101-129.

the requirement $L_\rho(\gamma) \geq 1$, for all ρ and γ, in which case (4) takes the simple form

$$\lambda_D(E_1, E_2) = \sup_\rho 1/A_\rho(D), \qquad L_\rho(\gamma) \geq 1. \tag{4'}$$

The extremal distance is a *conformal invariant* of the configuration (D, E_1, E_2), i. e. if $z^* = f(z)$ is a one-to-one conformal mapping of D upon a region D^*, taking E_1 to E^*_1 and E_2 to E^*_2, then

$$\lambda_{D^*}(E^*_1, E^*_2) = \lambda_D(E_1, E_2). \tag{5}$$

In fact, with $\rho^*(z^*) = \rho(z)/|f'(z)|$ we have

$$L_{\rho^*}(\gamma^*) = \int_{\gamma^*} \rho^* \, |dz^*| = \int_\gamma \rho \, |dz| = L_\rho(\gamma),$$

$$A_{\rho^*}(D^*) = \iint_{D^*} \rho^{*2} dx^* dy^* = \iint_{D^*} \rho^2/|f'(z)|^2 dx^* dy^*$$
$$= \iint_D \rho^2 dx dy = A_\rho(D).$$

The definition of extremal distance remains meaningful if we allow E_1 and E_2 to contain accessible boundary points of D.

2. The extension principle. The following simple principle is an important tool in dealing with extremal distances: If D^* is a region containing D, and if $E_1{}^*$ and $E_2{}^*$ are compact subsets of D^* containing E_1 and E_2 respectively, then

$$\lambda_{D^*}(E_1{}^*, E_2{}^*) \leq \lambda_D(E_1, E_2). \tag{6}$$

The proof is immediate: Any $\rho(z)$ defined on D^* is also defined on D,

$$A_\rho(D) \leq A_\rho(D^*), \tag{7}$$

and, since any curve joining E_1 and E_2 will a fortiori belong to the class Γ^* of curves joining $E_1{}^*$ and $E_2{}^*$,

$$\inf L_\rho(\gamma) \geq \inf L_\rho(\gamma^*), \text{ where } \gamma \,\varepsilon\, \Gamma,\ \gamma^* \,\varepsilon\, \Gamma^*. \tag{8}$$

Inserted into the definition (4), these inequalities yield (6).

In this connection we observe that if $E_1{}'$ and $E_2{}'$ denote the boundaries [2] of E_1 and E_2, we have $E_1{}' \subset E_1$, $E_2{}' \subset E_2$; and (6) gives

$$\lambda_D(E_1, E_2) \leq \lambda_D(E_1{}', E_2{}'). \tag{9}$$

---

[2] Throughout this paper primed letters will denote the boundaries of the respective sets.

On the other hand every curve joining E_1 and E_2 will also join their boundaries, which implies, in analogy with (8), that

$$\lambda_D(E_1, E_2) \geq \lambda_D(E_1', E_2'). \tag{10}$$

Combining (9) and (10) we have

$$\lambda_D(E_1, E_2) = \lambda_D(E_1', E_2'). \tag{11}$$

i. e. for purposes of finding the extremal distance we may replace E_1 and E_2 by their boundaries.

3. Determination of a maximal ρ-function. With certain restrictions on E_1, E_2, and D, necessary for the application of the classical existence theorems on harmonic functions, we will now determine a function $\rho \,\varepsilon\, \mathrm{P}$ for which the quantity (3) actually attains a maximum. Suppose $D^* = D \cap C(E_1 \cup E_2)$ is a region whose boundary consists of a finite number of analytic curves. Then it is known that there exists a function $u(x, y)$ harmonic in D^* with the boundary values 1 on E_1', 0 on E_2' and normal derivative zero on D'. *We assert that*

$$\rho_0(z) = (u_x^2 + u_y^2)^{\frac{1}{2}} = |\operatorname{grad} u| \tag{12}$$

maximizes (3).

To prove this we observe that all but a finite number of the level curves of the conjugate harmonic function v of u must join E_1' and E_2'. In fact, grad u can vanish only at a finite number of points, and u is monotonic on any curve $v = \text{const.}$ not containing such a point, which implies that the curve cannot be closed nor have both endpoints on E_1' or both on E_2'. It can of course not continue endlessly in the interior of D^*, since the existence of an accumulation point would then yield $v \equiv \text{const.}$ Also, a level curve of v with grad $u \neq 0$ is disjoint from D', since $\partial u/\partial n$ was assumed to vanish there.

Now let $\rho(z)$ be any member of the class ρ, normalized as in (4′). Integrating along any level curve of v joining E_1' and E_2' we have

$$\int_{v=c} \rho/\rho_0 \, du = \int_{v=c} \rho/(\partial u/\partial s) du = \int_{v=c} \rho \, |dz| \geq 1 = \int_{v=c} du; \tag{13}$$

hence, since the number of exceptional curves is finite,

$$\iint_{D^*} \rho/\rho_0 \, dudv \geq \iint_{D^*} dudv. \tag{14}$$

But from this it follows that

$$\iint_{D^*} (\rho/\rho_0 - 1)^2 dudv \leq \iint_{D^*} \rho^2/\rho_0^2 \, dudv - \iint_{D^*} dudv, \tag{15}$$

i. e. expressed in x and y, ρ_0^2 being the Jacobian,

$$(16)\qquad 0 \leq \iint_{D^*} (\rho - \rho_0)^2 dxdy \leq A_\rho(D^*) - A_{\rho_0}(D^*).$$

On account of the form (4′) of the definition of extremal distance, and (11), this means that

$$(17)\qquad \lambda_D(E_1, E_2) = \lambda_{D^*}(E_1', E_2') = 1/A_{\rho_0}(D^*) = 1/\iint_{D^*}(u_x^2 + u_y^2)\,dxdy = 1/D(u).$$

Here and below $D(u)$ denotes the Dirichlet integral taken over

$$D^* = D \cap C(E_1 \cup E_2).$$

This completes the proof of (12).

We may observe that in the restricted case considered in this section, and actually by a suitable limit process in a more general case, the relation (17) could be used as the definition of extremal distance. In view of the electrostatic interpretation, λ could then be called the *resistance* between E_1 and E_2.

It is easy to extend the validity of (12) to the case where E_1, E_2, and D are still bounded by a finite number of analytic curves, but D^* is not connected. The set D^* will then be the sum of a finite number n of regions D_i. For each one of those components D_i, say $D_1, \cdots, D_m$, $m \leq n$, whose boundary contains both a part E_1^i of E_1 and a part E_2^i of E_2, there exists an extremal distance for which we obtain by (4′)

$$(4'')\qquad 1/\lambda_{D_i}(E_1, E_2) = 1/\lambda_{D_i}(E_1^i, E_2^i) = \inf_{\rho_i} A_{\rho_i}(D_i), \qquad L_{\rho_i}(\gamma_i) \geq 1.$$

The remaining $D_{m+1}, \cdots, D_n$ have no influence upon the extremal distance between E_1 and E_2, since no curve γ will pass through them. Setting $\rho = \rho_i$ for $i = 1, \cdots, m$, and $\rho \equiv 0$ for $i > m$, we have $L_\rho(\gamma) \geq 1$; hence [3]

$$(18)\qquad 1/\lambda_D(E_1, E_2) = \inf_\rho A_\rho(D) = \sum_{i=1}^m \inf_{\rho_i} A_{\rho_i}(D_i) = \sum_{i=1}^m 1/\lambda_{D_i}(E_1^i, E_2^i) = \sum_{i=1}^m 1/\lambda_D(E_1^i, E_2^i),$$

and by (17) in evident notation

$$(17')\qquad 1/\lambda_D(E_1, E_2) = \sum_{i=1}^m D(u_i) = D(u).$$

[3] The last member of equation (18) has been inserted only for future reference.

4. Continuity of λ. If the sets E_1 and E_2 do not have the simple structure assumed in section **3**, we cannot in general find a harmonic function in D^* with preassigned boundary values and hence the extremal ρ-function, if it exists in the general case, cannot be determined as above. To this end we will prove the following lemma, which enables us to extend many of the subsequent results on extremal distance to any compact sets.

LEMMA. *Let D be a region, and E and F two disjoint compact subsets of D. Then if $\{E_n, F_n\}$ is a sequence of compact subsets of D, covering E and F respectively and converging to E and F (i. e. given $\epsilon > 0$ there exists N such that for $n > N$ every point of E_n and F_n is within distance ϵ of some point of E and F, respectively), we have*

$$\lim_{n\to\infty} \lambda_D(E_n, F_n) = \lambda_D(E, F). \tag{20}$$

We consider the normalized definition (4′), and begin by proving[4] that for any given ρ, the condition $L_\rho(\gamma) \geq 1$ for all γ joining E and F implies that, given $\epsilon > 0$, for $n > n_\epsilon$,

$$L_\rho(\gamma') > 1 - \epsilon \tag{21}$$

for all curves, denoted by γ', joining E_n and F_n. Let z_0 be any point of E, let C_r be the circle $|z - z_0| = r$ for $0 < r \leq k$, where k is such that $C_k \subset D$, and $f(r) = \inf_\gamma L_\rho(\gamma)$ for γ joining F and C_r. We wish to prove that

$$a = \lim_{r\to 0} f(r) \geq 1. \tag{22}$$

From this (21) follows, since E is compact, and the argument can be reapplied to F.

Suppose $a < 1$. Since ρ is square integrable on D we see, by an application of Schwarz' inequality, that for any given $d > 0$ the set $S(d)$ of values of r for which

$$\int_{C_r} \rho \, | dz | < d, \tag{23}$$

has $r = 0$ as a point of accumulation. Hence we can select a sequence $\{r_n\}$, $n = 1, 2, \cdots$ decreasing to zero, $r_1 \leq k$, such that

$$\sum_{n=1}^{\infty} \int_{C_{r_n}} \rho \, | dz | < (1 - a)/2.$$

---

[4] I am indebted to Professor Beurling for an unpublished communication containing the argument that follows.

But, for $\bar{\gamma}$ joining $C_{r_{n+1}}$ and C_{r_n},

$$\inf L_\rho(\bar{\gamma}) \leq f(r_{n+1}) - f(r_n)$$

Hence a curve γ can be constructed, joining F and z_0, such that

$$\begin{aligned} L_\rho(\gamma) &\leq f(r_1) + \sum_{n=1}^{\infty} [f(r_{n+1}) - f(r_n)] + \sum_{n=1}^{\infty} \int_{C_{r_n}} \rho \, | \, dz \, | + (1-a)/4 \\ &\leq a + (1-a)/2 + (1-a)/4 = (a+3)/4 < 1. \end{aligned}$$

This contradiction proves (22).

To obtain (20) from (21) let us first assume that $\lambda(E, F)$ is finite. Given ϵ, $0 < \epsilon < \lambda(E, F)$, by (4′) there is a ρ for which

$$1/A_\rho(D) > \lambda(E, F) - \epsilon.$$

For this ρ, and $n > n_\epsilon$, by (21) there exist E_n, F_n such that $L_\rho(\gamma') > 1 - \epsilon$ for all γ'. Hence the function $\beta = \rho/(1-\epsilon)$ is one satisfying the normalization (4′) in evaluating $\lambda(E_n, F_n)$. We have

$$\lambda(E_n, F_n) \geq 1/A_\beta(D) = (1-\epsilon)^2/A_\rho(D) > (1-\epsilon)^2(\lambda(E, F) - \epsilon).$$

If $\lambda(E, F)$ is infinite, given $M > 2$, there is a ρ with $1/A_\rho(D) > M$, $L_\rho \geq 1$, and the analogous reasoning gives

$$\lambda(E_n, F_n) \geq 1/A_\sigma(D) > (1 - 1/M)^2 M, \text{ where } \sigma = \rho/(1 - 1/M).$$

This completes the proof of (20).

In the definition (4) of extremal distance we assumed the sets E_1 and E_2 to be closed. This facilitates the statements and proofs of certain results, but it should be pointed out that the restriction is unessential. The definition (4) remains meaningful for arbitrary bounded sets E_1, E_2 with disjoint closures, and *the extremal distance thus defined is equal to the extremal distance between the closures* $\bar{E}_1$, $\bar{E}_2$.

The reasoning which led to the extension principle (6) immediately gives us the inequality

$$\lambda(E_1, E_2) \geq \lambda(\bar{E}_1, \bar{E}_2). \tag{24}$$

To prove the opposite inequality we use the normalized definition (4′) and hence wish to show that, if Γ denotes the class of rectifiable curves γ joining E_1 and E_2, and $\bar{\Gamma}$ the class of rectifiable curves $\bar{\gamma}$ joining $\bar{E}_1$ and $\bar{E}_2$, then for each fixed ρ,

$$\inf_{\gamma \in \Gamma} L_\rho(\gamma) \leq \inf_{\bar{\gamma} \in \bar{\Gamma}} L_\rho(\bar{\gamma}). \tag{25}$$

Let $\bar{\gamma}$ be any curve in $\bar{\Gamma}$, its endpoints $z \varepsilon \bar{E}_1$ and $\zeta \varepsilon \bar{E}_2$. For each positive integer n, let $z_n \varepsilon E_1$ and $\zeta_n \varepsilon E_2$ be points such that $| z - z_n | < 2^{-n}$ and $| \zeta - \zeta_n | < 2^{-n}$. Denote by γ_n the curve composed of the polygonal line $z_n, z_{n+1}, z_{n+2}, \cdots$, the curve $\bar{\gamma}$, and the polygonal line $\cdots, \zeta_{n+2}, \zeta_{n+1}, \zeta_n$. This γ_n belongs to Γ, and $\lim L_\rho(\gamma_n) = L_\rho(\bar{\gamma})$ as $n \to \infty$, which proves (25); hence

$$\lambda(E_1, E_2) = \lambda(\bar{E}_1, \bar{E}_2). \tag{26}$$

II. Representation of the Extremal Distance in Terms of a Generalized Potential.

We will now derive a representation of the extremal distance,[5] which will in particular be useful for obtaining estimations.

Let the region D and the compact subsets E_1 and E_2 be bounded by a finite number of analytic curves. Let L be a straight line intersecting D, and denote by $\bar{z}$, $\bar{E}_1$ etc. the points and sets symmetric to z, E_1 etc. with respect to L. Consider the set $E = (E_1 \cap \bar{E}_1) \cup ((E_2 \cap \bar{E}_2)$. If we identify symmetric boundary points of E, a finite number of Riemann surfaces are formed. We are going to apply to the set $D \cap C(E_1 \cup E_2)$, which is now contained in the Riemann surfaces, certain methods similar to those of logarithmic potential theory in the plane. The reader may in a first reading wish to follow the argument in the plane case and may do so by assuming E to be empty or to be situated on the line L. For the applications in Chapter III, the general case is needed, however.

For simplicity we will denote any one of the Riemann surfaces constructed above by D, and by E_1, E_2 and E the intersections of the original sets E_1, E_2 and E with this D. By the existence theorem for abelian integrals of the third kind there is a function $G(\zeta, z_1, z_2)$ with the following properties: Given any two distinct points z_1, z_2 of D, the difference

$$G(\zeta, z_1, z_2) - \log (| \zeta - z_2 | / | \zeta - z_1 |) \tag{1}$$

is harmonic for $\zeta \varepsilon D$; when $\zeta \varepsilon E$, $G(\zeta, z_1, z_2) = G(\bar{\zeta}, z_1, z_2)$ and

$$\partial G(\zeta, z_1, z_2)/\partial n = - \partial G(\bar{\zeta}, z_1, z_2)/\partial n;$$

and $\partial G/\partial n = 0$ on the remaining boundary of D. Since G is only determined up to a constant, we normalize it by requiring the difference (1) to vanish

[5] The possibility of such a representation was suggested to me by Professor Ahlfors.

at an auxiliary point $\zeta = z_0$ on $L \cup E$. For later use we observe that the relations

$$G(\zeta, z_1, z_2) + G(\zeta, z_2, z_1) = 0. \tag{2}$$

$$G(\zeta, z_1, z_2) + G(\zeta, z_2, z_3) + G(\zeta, z_3, z_1) = 0, \tag{3}$$

hold for $z_3 \varepsilon D$, $z_3 \neq z_1, z_2$. In fact, let $u(\zeta)$ denote for a moment the left member of either (2) or (3); u is harmonic throughout D, $u(z_0) = 0$ and the Dirichlet integral $\int_{D'} u(\partial u/\partial n) \mid d\zeta \mid$ vanishes due to the properties of G on the boundary D' of D. Hence u is identically zero. To see that $G(\zeta, z_1, z_2)$ is harmonic also in z_1 and z_2 we choose a point z_4 in D, distinct from z_1, z_2 and z_3, and conclude from

$$\int_{D'} [G(\zeta, z_1, z_2) \partial G(\zeta, z_3, z_4)/\partial n - G(\zeta, z_3, z_4) \partial G(\zeta, z_1, z_2)/\partial n] \mid d\zeta \mid = 0$$

that

$$G(z_1, z_3, z_4) - G(z_2, z_3, z_4) - G(z_3, z_1, z_2) + G(z_4, z_1, z_2) = 0 \tag{4}$$

Let M_i be the set of all Borel distributions μ_i on the boundary E_i' of E_i with $\mu_i(E_i') = 1$, $i = 1, 2$, i. e. measures for which every open set is measurable. For any $z \varepsilon D$, the abstract Lebesgue integral of $\log \mid \zeta - z \mid$ over E_i' with respect to any such unit distribution is well defined. Thus we may consider the function

$$p(z_1, z_2) = \int_{E_1'} G(\zeta, z_1, z_2) \, d\mu_1(\zeta) - \int_{E_2'} G(\zeta, z_1, z_2) \, d\mu_2(\zeta) \tag{5}$$

for any $\mu_1 \varepsilon M_1$, $\mu_2 \varepsilon M_2$ and all values of z_1 and z_2 in D except those for which both points fall simultaneously on E_1' or simultaneously on E_2'. Differentiating under the integral signs we find $p(z_1, z_2)$ to be harmonic in both variables in $D \cap C(E_1' \cup E_2')$. When $z_1 \varepsilon E_1'$ or $z_2 \varepsilon E_2'$ or both, $p(z_1, z_2)$ is lower semi-continuous. To see this we may consider the truncated functions

$$p_n(z_1, z_2) = \int_{E_1'} \min[G(\zeta, z_1, z_2), n] \, d\mu_1(\zeta) - \int_{E_2'} \max[G(\zeta, z_1, z_2), -n] \, d\mu_2(\zeta),$$

which are continuous and increase to $p(z_1, z_2)$ (which may be $+\infty$) as $n \to \infty$. Analogously, if $z_1 \varepsilon E_2'$ or $z_2 \varepsilon E_1'$ or both, we find $p(z_1, z_2)$ to be upper semi-continuous. In particular we conclude that, for each fixed pair of unit distributions, the corresponding function $p(z_1, z_2)$ will attain a minimum for $z_1 \varepsilon E_1'$ and $z_2 \varepsilon E_2'$.

The main theorem of this chapter can now be expressed as follows:

THEOREM. *Let the region D, on a Riemann surface, and the disjoint compact subsets E_1 and E_2 of D be bounded by a finite number of analytic curves. For each pair μ_1, μ_2 of Borel unit distributions on the boundaries E_1', E_2' of E_1, E_2, let the function $p(z_1, z_2)$ be defined by* (5), *where the kernel $G(\zeta, z_1, z_2)$ is defined by* (1). *Then there exists among these pairs of distributions a pair for which the quantity*

$$d = d(\mu_1, \mu_2) = \min_{z_i \varepsilon E_i'} p(z_1, z_2)$$

is maximal, and this maximum is equal to 2π times the extremal distance between E_1 and E_2:

$$\lambda_D(E_1, E_2) = 1/2\pi \max_{\mu_i \varepsilon M_i} d(\mu_1, \mu_2) = 1/2\pi \max_{\mu_i \varepsilon M_i} \min_{z_i \varepsilon E_i'} p(z_1, z_2). \tag{6}$$

To prove (6) we first wish to show that

$$\lambda(E_1, E_2) \geqq 1/2\pi\, d(\mu_1, \mu_2) \tag{7}$$

for all $\mu_i \varepsilon M_i$, and then construct the extremal distributions. Given any fixed pair $\mu_i \varepsilon M_i$, it is possible to fix z_2 on E_2' so that $p(z, z_2)$ is still defined for $z \varepsilon E_2'$ and non-positive there. In fact, if z_3 is any fixed point in $D \cap C(E_2')$ we can, by the upper semicontinuity of p, choose z_2 to be a point of E_2' at which $p(z, z_3)$ attains its maximum; since by (3) and (2)

$$p(z, z_2) = p(z, z_3) - p(z_2, z_3), \tag{8}$$

the desired non-positiveness follows. Given a positive number ϵ smaller than $d/2$, denote by E_1^* the set where $p(z, z_2) \geqq d - \epsilon$, and by E_2^* the set where $p(z, z_2) \leq \epsilon$. The set E_1^* contains E_1' by the definition of d, and E_2^* contains E_2' by our choice of z_2. By the lower semi-continuity of p on E_1' and the upper semi-continuity on E_2', respectively, the boundaries $E_1^{*\prime}$ and $E_2^{*\prime}$ of E_1^* and E_2^* are disjoint from E_1' and E_2'; hence, as level curves of a harmonic function, they are composed of a finite number of analytic curves. The function

$$p^*(z) = (p(z, z_2) - \epsilon)/(d - 2\epsilon), \tag{9}$$

harmonic in $D \cap C(E_1^* \cup E_2^*)$, has boundary values 0 on $E_1^{*\prime}$ and 1 on $E_2^{*\prime}$, and normal derivative 0 on the remaining boundary. By Ch. I, sec. 3, we conclude that

$$\lambda(E_1^*, E_2^*) = 1/D(p^*) = (d - 2\epsilon)^2/D(p). \tag{10}$$

But for any c, $0 < c < d$, we have, n denoting the inner normal,

5

$$(11)\quad D(p) = -d\int_{p=c}\partial p/\partial n\,|dz| = -d\int_{p=c}|dz|\int_{E_1'}\partial G(\zeta_1, z, z_2)/\partial n\, d\mu_1(\zeta_1)$$
$$+\, d\int_{p=c}|dz|\int_{E_2'}\partial G(\zeta_2, z, z_2)/\partial n\, d\mu_2(\zeta_2)$$
$$= -d\int_{E_1'}d\mu_1(\zeta_1)\int_{p=c}\partial G(\zeta_1, z, z_2)/\partial n\,|dz|$$
$$+\, d\int_{E_2'}d\mu_2(\zeta_2)\int_{p=c}\partial G(\zeta_2, z, z_2)/\partial n\,|dz|$$
$$= d\int_{E_1'}d\mu_1(\zeta_1)\int_{E_2'}d\mu_2(\zeta_2)\int_{p=c}[\partial G(\zeta_2, z, z_2)/\partial n - \partial G(\zeta_1, z, z_2)/\partial n]\,|dz|$$
$$= -d\int_{E_1'}d\mu_1(\zeta_1)\int_{E_2'}d\mu_2(\zeta_2)\int_{p=c}\partial G(z, \zeta_1, \zeta_2)/\partial n\,|dz| = 2\pi d,$$

where the change of the order of integration is justified by Fubini's theorem, the next to the last step is a consequence of (4), and the last equality is verified separately for each of the topologically different mutual positions of the sets E_1, E_2 and the points ζ_1, ζ_2 on them. Substituting (11) into (10) we have, on account of the extension principle (6), Chapter I,

$$\lambda(E_1, E_2) \geq \lambda(E_1^*, E_2^*) = (d - 2\epsilon)^2/2\pi d,$$

hence,

$$(7)\qquad \lambda(E_1, E_2) \geq d/2\pi.$$

To find distributions for which equality holds in (7), we consider the function $u(\zeta)$ which is harmonic on the open set $D^* = D \cap C(E_1 \cup E_2)$, takes the boundary values 0 on E_2' and $\lambda(E_1, E_2)$ on E_1', and has normal derivative zero on D'. Denoting by n the inner normal, we define for all subsets e of E_1' and E_2' respectively,

$$(12)\qquad \begin{aligned}\mu_1(e) &= \int_e |\partial u/\partial n|\,|d\zeta|\\ \mu_2(e) &= \int_e |\partial u/\partial n|\,|d\zeta|.\end{aligned}$$

By (17), Ch. I, for $i = 1, 2$,

$$\int_{E_i'}|\partial u/\partial n|\,|d\zeta| = -1/\lambda\int_{E_1'}u(\partial u/\partial n)\,|d\zeta| = D(u)/\lambda$$
$$= \lambda^2 D(u/\lambda)/\lambda = 1.$$

We will show that for the corresponding function p the relation

$$(13)\qquad p(z_1, z_2) = -\int_{E_1' \cup E_2'}G(\zeta, z_1, z_2)\partial u(\zeta)/\partial n\,|d\zeta| = 2\pi\cdot\lambda(E_1, E_2)$$

holds for all $z_1 \varepsilon E_1'$, $z_2 \varepsilon E_2'$, which in particular together with (7) will imply (6).

First let us assume z_1 and z_2 to be interior to D^*. If C_i are circles in D^* with centers z_i and radii r, we apply Green's formula in each component of D^* and obtain by adding,

$$(14) \qquad 1/2\pi \int_{D' \cup E_1' \cup E_2' \cup C_1 \cup C_2} (G\partial u/\partial n - u\partial G/\partial n) \, | \, d\zeta \, | = 0$$

which for $r \to 0$ takes the form

$$(15) \qquad u(z_1) - u(z_2) = -1/2\pi \int_{E_1' \cup E_2'} G\partial u/\partial n \, | d\zeta \, | = 1/2\pi \quad p(z_1, z_2).$$

To extend the validity of (15) to the case where z_i are on E_i' and thus complete the proof of (13), we only have to observe that $p(z_1, z_2)$ is continuous in z_1 and z_2 on the closure of D^*. This follows from the fact that $p(z_1, z_2)$ differs by a continuous function from the ordinary logarithmic potentials of the given distributions with continuous densities.[6]

III. Estimations by Transportation and Projection.

1. The symmetrization theorem for the entire plane. The original definition (4), Chapter I, includes the simplest lower estimation of $\lambda_D(E_1, E_2)$: If we choose a particular function $\rho(z)$, the value of the quantity (3) will not exceed λ_D.

The extension principle (6), Chapter I, yields both an upper and a lower estimation: We may construct within E_1 and E_2 sets whose extremal distance we are able to compute explicitly, and we may cover E_1 and E_2 by sets with this property.

In this chapter we shall show how the representation (6), Chapter II, enables us to find interesting upper estimations by deforming and moving the sets E_i, e. g. placing them in symmetric positions with respect to a given axis or projecting them upon a given curve. We commence with the following

THEOREM.[7] *Let E_1 and E_2 be two disjoint compact subsets of the z-plane, D. For every $r \geq 0$ denote by C_r the circle $| z | = r$, and by $\alpha_1(r)$ and $\alpha_2(r)$*

---

[6] See O. D. Kellogg, *Foundations of Potential Theory*, 1929, Chapter VI.

[7] A theorem of the same nature as this for another type of circular symmetrization is given, in the case where the sets are bounded by analytic curves, by G. Pólya, *Comptes Rendus*, Paris, 1950, t. 230, p. 25. Also compare G. Pólya and G. Szegö, *American Journal of Mathematics*, vol. 67 (1945), p. 1-32. Their proof is based on an entirely different idea.

the angular Lebesgue measures of the sets $E_1 \cap C_r$ *and* $E_2 \cap C_r$*, respectively. If the sets* $\tilde{E}_1$ *and* $\tilde{E}_2$ *are defined by the inequalities*

$$(1)\qquad \begin{aligned} \tilde{E}_1:&\quad \pi - \tfrac{1}{2}\alpha_1(r) \leq \phi(r) \leq \tfrac{1}{2}\alpha_1(r) + \pi, \\ \tilde{E}_2:&\quad -\tfrac{1}{2}\alpha_2(r) \leq \phi(r) \leq \tfrac{1}{2}\alpha_2(r), \end{aligned}$$

ϕ *and* r *being polar coordinates in* D*, then* [8]

$$(2)\qquad \lambda_D(E_1, E_2) \leq \lambda_D(\tilde{E}_1, \tilde{E}_2).$$

Our proof of this theorem rests essentially upon

LEMMA 1. *Let* E_1 *and* E_2 *be two disjoint compact sets with boundaries composed of a finite number of analytic curves. Let* L *be a directed straight line through the origin; denote the half plane to the right of* L *by* H*, the left by* $\bar{H}$*. If* A *is any set in the plane, denote by* $\bar{A}$ *the set symmetric to* A *with respect to* L*. Define the sets* $E_1{}^*$ *and* $E_2{}^*$ *as follows:* [9]

$$(3)\qquad \begin{aligned} E_1{}^* &= (E_1 \cap \bar{E}_1) \cup (E_1 \cap \bar{H}) \cup (\bar{E}_1 \cap \bar{H}) \\ E_2{}^* &= (E_2 \cap \bar{E}_2) \cup (E_2 \cap H) \cup (\bar{E}_2 \cap H). \end{aligned}$$

Then

$$(4)\qquad \lambda_D(E_1, E_2) \leq \lambda_D(E_1{}^*, E_2{}^*),$$

where D *again is the entire plane.*[10]

We begin by observing that the reasoning leading to the formula (18), Ch. I, remains valid in the case where the D_i are the Riemann surfaces intro-

---

[8] It is not obvious that the sets $\tilde{E}_1$ and $\tilde{E}_2$ are closed. Those who wish, in defining extremal distance, to restrict themselves to closed sets, may prove the closedness as follows:

Consider a sequence $r_n e^{i\phi_n}$ of points of $\tilde{E}_1$, converging to a point $r_0 e^{i\phi_0}$. It follows from the definition of $\tilde{E}_1$ that the corresponding sequence of sets $E_1 \cap C_{r_n}$ has the property that $\liminf \alpha_1(r_n) \geq 2\phi_0$. We wish to prove that this implies $\alpha_1(r_0) \geq 2\phi_0$. Since E_1 is closed, $\alpha_1(r_0)$ is greater than or equal to the measure β of the set of points on $E_1 \cap C_{r_0}$ which are limit points of sequences of points of the sets $E_1 \cap C_{r_n}$. But it is well-known (cf. E. J. McShane, *Integration*, p. 105) that $\beta \geq \limsup \alpha_1(r_n) \geq 2\phi_0$.

A similar remark applies to the theorems in section 2 of this chapter.

Incidentally, the closedness of the original sets E_1 and E_2 in the above theorem was used to assure the measurability of their intersections with the circles C_r.

[9] In words: To obtain E^*_1 from E_1 we replace by their symmetric images in $\bar{H}$ those parts of $E_1 \cap H$ whose symmetric images do not belong to E_1. Similarly, we move the parts of $E_2 \cap \bar{H}$ whose images do not belong to E_2.

[10] The principle underlying this approach is related to the method of "rearrangement" discussed by Hardy, Littlewood and Pólya in their *Inequalities*, Cambridge, 1934.

duced at the beginning of Ch. II. Hence it is sufficient to prove Lemma 1 for one such Riemann surface. As before, we will denote it by D, and the boundaries of $E_1 \cap D$ and $E_2 \cap D$ by E_1' and E_2', respectively.

Consider an arbitrary fixed pair of unit distributions μ_1, μ_2 on E_1' and E_2'. Define a corresponding pair μ_1^*, μ_2^* on $E_1^{*\prime}, E_2^{*\prime}$ as follows: [11]

$$(5)\quad \begin{aligned} \mu_1^*(A) &= \mu_1(A) \text{ for } A \subset (E_1' \cap \bar{E}_1') \cup (E_1' \cap \bar{H}), \\ \mu_1^*(A) &= \mu_1(\bar{A}) \text{ for } A \subset (\bar{E}_1' \cap \bar{H}) \cap C(E_1' \cap \bar{E}_1'), \\ \mu_2^*(A) &= \mu_2(A) \text{ for } A \subset (E_2' \cap \bar{E}_2') \cup (E_2' \cap H), \\ \mu_2^*(A) &= \mu_2(\bar{A}) \text{ for } A \subset (\bar{E}_2' \cap H) \cap C(E_2' \cap \bar{E}_2'). \end{aligned}$$

Similarly, if $z_1 \varepsilon E_1'$, $z_2 \varepsilon E_2'$ is any pair of points on the original sets we define

$$(6)\quad \begin{aligned} z_1^* &= z_1 \text{ for } z_1 \varepsilon (E_1' \cap \bar{E}_1') \cup (E_1' \cap \bar{H}), \\ z_1^* &= \bar{z}_1 \text{ for } z_1 \varepsilon (E_1' \cap H) \cap C(E_1' \cap \bar{E}_1'), \\ z_2^* &= z_2 \text{ for } z_2 \varepsilon (E_2' \cap \bar{E}_2') \cup (E_2' \cap H), \\ z_2^* &= \bar{z}_2 \text{ for } z_2 \varepsilon (E_2' \cap \bar{H}) \cap C(E_2' \cap \bar{E}_2'). \end{aligned}$$

Setting

$$(7)\quad \int_{E_1^{*\prime}} G(\zeta, z_1, z_2)\, d\mu_1^*(\zeta) - \int_{E_2^{*\prime}} G(\zeta, z_1, z_2)\, d\mu_2^*(\zeta) = p^*(z_1, z_2),$$

we wish to show that

$$(8)\quad p(z_1, z_2) \leq p^*(z_1^*, z_2^*).$$

Since (6) sets up a one-to-one correspondence between the pairs of points on E_1', E_2' and $E_1^{*\prime}, E_2^{*\prime}$, (8) implies that

$$(9)\quad \min_{z_i \varepsilon E_i'} p(z_1, z_2) \leq \min_{z_i \varepsilon E_i^{*\prime}} p^*(z_1, z_2)$$

for each pair of distributions; hence by (6), Ch. II, and (11), Ch. I,

$$(10)\quad \lambda_D(E_1, E_2) \leq \lambda_D(E_1^*, E_2^*).$$

We shall collect here a few properties of the function G, which we will need in the verification of (8): If z_1 and z_2 are interior points of $D \cap C(E)$ (we recall that $E = (E_1 \cap \bar{E}_1) \cup (E_2 \cap \bar{E}_2))$, the function

$$(11)\quad G(\bar{\zeta}, z_1, z_2) - G(\zeta, z_1, z_2)$$

---

[11] In words: if the set A moves in the transformation (3), the mass on A moves with it.

is a harmonic function of ζ in D except for positive poles at $\bar{z}_1$ and z_2 and negative poles at z_1 and $\bar{z}_2$, and vanishes on $E \cup L$. Hence it is non-positive if ζ and z_1 are on the same side of L and z_2 on the opposite side, and non-negative if ζ and z_2 are on the same side of L and z_1 on the opposite side. If z_1 is a point of E, the function (11) is singular only at z_2 and $\bar{z}_2$ (since on the Riemann surface the opposite poles at z_1 and $\bar{z}_1$ coincide), hence non-negative as soon as ζ and z_2 are on the same side of L and non-positive if they are on opposite sides; and analogously for $z_2 \varepsilon E$. If both $z_1 \varepsilon E$ and $z_2 \varepsilon E$, the function (11) vanishes identically.

If z_1 is an interior point and z_2 either an interior point or a point of E, the function

$$G(\zeta, \bar{z}_1, z_2) - G(\zeta, z_1, z_2) \tag{12}$$

is, by (3) and (2), Ch. II, equal to $G(\zeta, \bar{z}_1, z_1)$. On $E \cup L$,

$$G(\zeta, \bar{z}_1, z_1) = G(\bar{\zeta}, \bar{z}_1, z_1) = G(\zeta, z_1, \bar{z}_1) = -G(\zeta, \bar{z}_1, z_1)$$

(the middle equality follows from the fact that $G(\bar{\zeta}, \bar{z}_1, z_1) - G(\zeta, z_1, \bar{z}_1)$ is harmonic everywhere and vanishes at z_0), i.e. $G(\zeta, \bar{z}_1, z) = 0$. Hence the function (12) is non-negative when ζ is on the same side of L as $\bar{z}_1$ and non-positive on the opposite side. If $z_1 \varepsilon E$, (12) is identically zero. Analogous considerations apply to the function

$$G(\zeta, z_1, \bar{z}_2) - G(\zeta, z_1, z_2). \tag{13}$$

The verification of (8) must be carried out separately for the different possibilities arising from (6). First let us assume that $z_1^* = z_1$, $z_2^* = z_2$; then

$$\begin{aligned}
p^*(z_1^*, z_2^*) - p(z_1, z_2) &= \int_{\bar{E}_1' \cap \bar{H} \cap C(E)} G(\zeta, z_1, z_2)\,d\mu_1(\zeta) \\
&\quad - \int_{E_1' \cap H \cap C(E)} G(\zeta, z_1, z_2)\,d\mu_1(\zeta) - \int_{\bar{E}_2' \cap H \cap C(E)} G(\zeta, z_1, z_2)\,d\mu_2(\zeta) \\
&\quad + \int_{E_2' \cap \bar{H} \cap C(E)} G(\zeta, z_1, z_2)\,d\mu_2(\zeta) \\
&= \int_{E_1' \cap H \cap C(E)} [G(\bar{\zeta}, z_1, z_2) - G(\zeta, z_1, z_2)]\,d\mu_1(\zeta) \\
&\quad - \int_{E_2' \cap \bar{H} \cap C(E)} [G(\bar{\zeta}, z_1, z_2) - G(\zeta, z_1, z_2)]\,d\mu_2(\zeta).
\end{aligned} \tag{14}$$

This is seen to be non-negative by the properties of the function (11) discussed above; e.g. in the first integral on the right either $z_1 \varepsilon \bar{H} \cap C(E)$,

$z_2 \varepsilon H \cap C(E)$, in which case ζ and z_2 are on the same side and z_1 on the opposite side of L, or one of the points, e. g. z_1, is on E, in which case ζ and z_2 are on the same side of L, or both points are on E. Next suppose $z_1^* = \bar{z}_1$, $z_2^* = z_2$; then

$$(15)\quad p^*(z_1^*, z_2^*) - p(z_1, z_2) = \int_{E_1' \cap H \cap C(E)} [G(\bar{\zeta}, \bar{z}_1, z_2) - G(\zeta, z_1, z_2)] d\mu_1(\zeta)$$

$$+ \int_{(E_1' \cap \bar{H}) \cup (E_1' \cap \bar{E}_1')} [G(\zeta, \bar{z}_1, z_2) - G(\zeta, z_1, z_2)] d\mu_1(\zeta)$$

$$- \int_{(E_2' \cap H) \cup (E_2' \cap \bar{E}_2')} [G(\zeta, \bar{z}_1, z_2) - G(\zeta, z_1, z_2)] d\mu_2(\zeta)$$

$$- \int_{E_2' \cap \bar{H} \cap C(E)} [G(\bar{\zeta}, \bar{z}_1, z_2) - G(\zeta, z_1, z_2)] d\mu_2(\zeta) \geq 0$$

by the properties of (12) and (13), observing the equation

$$G(\bar{\zeta}, \bar{z}_1, z_2) = G(\zeta, z_1, \bar{z}_2).$$

The case $z_1^* = z_1$, $z_2^* = \bar{z}_2$ is analogous to (15).· Finally suppose $z_1^* = \bar{z}_1$, $z_2^* = \bar{z}_2$; then

$$(16)\quad p^*(z_1^*, z_2^*) - p(z_1, z_2) = \int_{(E_1' \cap \bar{H}) \cup (E_1' \cap \bar{E}_1')} [G(\zeta, \bar{z}_1, \bar{z}_2) - G(\zeta, z_1, z_2)] d\mu_1(\zeta)$$

$$- \int_{(E_2' \cap H) \cup (E_2' \cap \bar{E}_2')} [G(\zeta, \bar{z}_1, \bar{z}_2) - G(\zeta, z_1, z_2)] d\mu_2(\zeta) \geq 0$$

by (11), since

$$G(\zeta, \bar{z}_1, \bar{z}_2) = G(\bar{\zeta}, z_1, z_2).$$

The inequality (8) is verified, and lemma 1 proved. We proceed to

LEMMA 2. *Let E_1 and E_2 be two sets of the type defined by the following inequalities:*

$$(17)\qquad \left.\begin{array}{l} (\nu - 1)\delta \leq r \leq \nu\delta \\ \phi^\nu_{i-1} \leq \phi \leq \phi_i^\nu, \quad i = 1, \cdots, m_\nu \end{array}\right\} \nu = 1, \cdots, n,$$

where ϕ_i^ν are numbers in the interval $0 \leq \phi \leq 2\pi$, and $\delta > 0$ (i. e. sets composed by a finite number of "concentric rectangles" with altitude δ). Then E_1 and E_2 can, by a finite sequence of transformations each of which does not decrease the extremal distance, be transformed into their corresponding symmetric images defined by (1).

First we consider, for a fixed integer ν, the subset E_1^ν of E_1 consisting of m_ν rectangles in the annulus $(\nu - 1)\delta \leq r \leq \nu\delta$. We observe that in a

transformation (3) in any line L through the origin the number m_ν of disjoint rectangles remains the same. A point z of $E_1{}^{\nu *}$ not in $E_1{}^\nu$ belongs to $(\bar{E}_1{}^\nu \cap \bar{H}) \cap C(E_1{}^\nu \cap \bar{E}_1{}^\nu)$; either it is connected with the set $E_1{}^\nu \cap \bar{E}_1{}^\nu$, in which case $\bar{z}$ was connected with $E_1{}^\nu \cap \bar{E}_1{}^\nu$ before the transformation, or it belongs to a rectangle R not connected with $E_1{}^\nu \cap \bar{E}_1{}^\nu$, in which case $\bar{R}$ was a disjoint rectangle of $E_1{}^\nu$. The same reasoning applies to points of $E_1{}^\nu$ not in $E_1{}^{\nu *}$. Now if $\phi_1{}^\nu \leq \phi \leq \phi_2{}^\nu$ and $\phi^\nu{}_3 \leq \phi \leq \phi_4{}^\nu$, with $0 \leq \phi_1{}^\nu \leq \phi_2{}^\nu \leq \phi_3{}^\nu \leq \phi_4{}^\nu$, are two rectangles of $E_1{}^\nu$, a transformation (3) in the axis $L: \frac{1}{2}(\phi_1{}^\nu + \phi_3{}^\nu)$ is seen to replace them by a rectangle of length $\phi_4{}^\nu - \phi_3{}^\nu + \phi_2{}^\nu - \phi_1{}^\nu$ and a line segment $\phi = \phi_1{}^\nu$, $(\nu - 1)\delta \leq r \leq \nu\delta$. By Lemma 1 the extremal distance does not decrease. Now we remove the line segment; by (6), Ch. I, the extremal distance does not decrease. Repeating this process $m_\nu - 1$ times we have transformed $E_1{}^\nu$ into a single rectangle with angular measure equal to that of $E_1{}^\nu$.

Having done this for all ν in both E_1 and E_2, we wish to transform the sets obtained, say $\hat{E}_1$, $\hat{E}_2$, into their symmetric images (1). Let R_1 be a rectangle of $\hat{E}_1$, the ϕ-coordinate of its center ϕ_1, with $-\pi < \phi_1 < \pi$. Then a transformation (3) in the axis $L: \frac{1}{2}(\phi_1 + \pi)$ will place R_1 symmetrically with respect to the negative real axis. For $R_2 \subset \hat{E}_2$ with center ϕ_2, $0 < \phi_2 < 2\pi$, a transformation (3) in the axis $L: \frac{1}{2}\phi_2$ places R_2 symmetrically with respect to the positive real axis. Using these transformations we can now move the components of $\hat{E}_1$ and $\hat{E}_2$, one at a time, into their symmetric positions. The fact that, by the definition (3), once a rectangle has reached its symmetric position it remains there under any transformation (3) in an axis between 0 and π—and only such axes L are used in this paragraph—completes the proof of Lemma 2.

The symmetrization theorem (2) can now be proved at once: Let the two disjoint compact sets E_1 and E_2 and positive numbers ϵ and M be given. Let the integer R be the radius of a circle with center at the origin, containing E_1 and E_2. For each integer n we consider the set of closed rectangles formed by the circles and rays

$$\begin{aligned} &r = \nu/2^n, && \nu = 0, 1, \cdots, 2^n R, \\ &\phi = \mu/2^n R, && \mu = 0, 1, \cdots, [2^{n+1}\pi R]. \end{aligned} \tag{18}$$

Denote by $E_1{}^n$ and $E_2{}^n$ the unions of those rectangles which contain points of E_1 and E_2, respectively. For n large enough, $E_1{}^n$ and $E_2{}^n$ will be disjoint, and by (20), Ch. I, n can be chosen so large that

$$\lambda(E_1{}^n, E_2{}^n) > \lambda(E_1, E_2) - \epsilon, \tag{19}$$

if $\lambda(E_1, E_2)$ is finite, and

$$\lambda(E_1{}^n, E_2{}^n) > M \tag{19'}$$

if $\lambda(E_1, E_2)$ is infinite. The symmetric images $\tilde{E}_1{}^n$, $\tilde{E}_2{}^n$ (see (1)) of $E_1{}^n$ and $E_2{}^n$ contain $\tilde{E}_1$ and $\tilde{E}_2$, respectively. Hence, by (6), Ch. I, and Lemma 2,

$$\lambda(\tilde{E}_1, \tilde{E}_2) \geq \lambda(\tilde{E}_1{}^n, \tilde{E}_2{}^n) \geq \lambda(E_1{}^n, E_2{}^n) > \lambda(E_1, E_2) - \epsilon, \tag{20}$$

in the finite case, or

$$\lambda(\tilde{E}_1, \tilde{E}_2) > M \tag{20'}$$

in the infinite case, i. e.

$$\lambda(\tilde{E}_1, \tilde{E}_2) \geq \lambda(E_1, E_2), \qquad \text{q. e. d.} \tag{2}$$

The extension principle (6), Ch. I, combined with (2) yields the following

Corollary. *Let E_1'' be the radial projection of E_1 upon the negative real axis: $-r \varepsilon E_1''$ if $C_r \cap E_1$ is not empty, and E_2'' the corresponding projection of E_2 upon the positive real axis: $r \varepsilon E_2''$ if $C_r \cap E_2$ is not empty. Then*

$$\lambda_D(E_1'', E_2'') \geq \lambda_D(\tilde{E}_1, \tilde{E}_2) \geq \lambda_D(E_1, E_2), \tag{21}$$

where D is the entire plane.

2. Similar results for other regions. To obtain results similar to the symmetrization theorem in cases where D is not the entire plane, we may utilize a simple reflexion principle exemplified by the proofs of the following theorems.

Theorem. *Let E_1 and E_2 be two closed sets in the unit circle D, disjoint from each other and from the origin. If $\tilde{E}_1$ and $\tilde{E}_2$ denote the corresponding sets defined by* (1), *then*

$$\lambda_D(E_1, E_2) \leq \lambda_D(\tilde{E}_1, \tilde{E}_2). \tag{22}$$

For the proof, let us first assume E_1 and E_2 to be bounded by a finite number of analytic curves. Let $\hat{E}_1$ and $\hat{E}_2$ be the images of E_1 and E_2 in a reflexion in the boundary C of D. Let $u(z)$ be the harmonic function in $D \cap C(E_1 \cup E_2)$ taking the value 1 on E_1, 0 on E_2, and with normal derivative 0 on C. By (17), Ch. I, we have

$$\lambda_D(E_1, E_2) = 1/D_D(u), \tag{23}$$

where $D_D(u)$ denotes the Dirichlet integral of u over $D \cap C(E_1 \cup E_2)$. In

the reflexion, $u(z)$ is extended into a function $u(z)$ with boundary values 1 on $E_1 \cup \hat{E}_1$, 0 on $E_2 \cup \hat{E}_2$, and harmonic in the entire exterior of these sets. Denoting the entire plane by D^*, we then have

$$(24) \quad \lambda_{D^*}(E_1 \cup \hat{E}_1, E_2 \cup \hat{E}_2) = 1/D_{D^*}(u) = 1/2D_D(u) = \tfrac{1}{2}\lambda_D(E_1, E_2).$$

Analogously,

$$(25) \qquad \lambda_{D^*}(\hat{E}_1 \cup \tilde{\hat{E}}_1, \hat{E}_2 \cup \tilde{\hat{E}}_2) = \tfrac{1}{2}\lambda_D(\hat{E}_1, \hat{E}_2).$$

But the symmetrization theorem in D^* applies to the left members of these equations; hence (22) follows for the quantities at right.

In the case where E_1 and E_2 are any disjoint closed sets not containing the origin, we cover them by rectangular sets and reason as in (20).

THEOREM. *Let E_1 and E_2 be two disjoint compact subsets of the set D:*

$$(26) \qquad 0 < \phi < \phi_0 < 2\pi, \qquad r > 0, \qquad z = re^{i\phi}.$$

For all $r > 0$ let C_r denote the circle $|z| = r$, and $\alpha_1(r)$ and $\alpha_2(r)$ the angular Lebesgue measures of the sets $E_1 \cap C_r$ and $E_2 \cap C_r$, respectively. If the sets E_1^* and E_2^* are defined by the inequalities

$$(27) \qquad \begin{aligned} &E_1^*\colon \quad 0 \leq \phi(r) \leq \alpha_1(r) \\ &E_2^*\colon \quad \phi_0 - \alpha_2(r) \leq \phi(r) \leq \phi_0, \end{aligned}$$

then

$$(28) \qquad \lambda_D(E_1, E_2) \leq \lambda_D(E_1^*, E_2^*).$$

Let us first assume E_1 and E_2 to be bounded by a finite number of analytic curves. If $\phi_0 \neq \pi$, we begin the proof by mapping D upon the upper half plane by the function $z'' = z^{\pi/\phi_0}$. By (5), Ch. I, the sets E_1, E_2, E_1^*, E_2^* go into sets E_1'', E_2'', $E_1^{*''}$, $E_2^{*''}$ with the same mutual extremal distances. Now a reflexion in the real axis yields sets $E_1'' \cup \hat{E}_1''$, etc. to which the symmetrization theorem for the plane applies. The relations (24) and (25) are again valid; we obtain our theorem for the half plane, and by the inverse mapping $z = (z'')^{\phi_0/\pi}$ for our original region D. The extension to arbitrary compact sets is performed as before.

The case where D is a half circle or more generally a circular sector is reduced to the above by a reflexion in the circular part of the boundary.

Extensions of the symmetrization theorem along another line are obtained simply by mapping one of the regions D of the above theorems conformally upon the region for which a symmetrization theorem is desired. Then, however, the symmetrization will not in general remain radial and euclidean.

3. Logarithmic capacity. Let E be a compact set bounded by a finite number of regular curves. Green's function $g(z)$ for the complement of E is uniquely determined by the following requirements: It is continuous for all z, harmonic in $C(E)$, zero on E, and

$$g(z) = \log |z| + c + \epsilon(|z|), \tag{35}$$

where $\epsilon(|z|) \to 0$ as $|z| \to \infty$. The *logarithmic capacity* (transfinite diameter, outer radius) of E is

$$\operatorname{Cap}(E) = e^{-c}. \tag{36}$$

In this section we will derive and apply a relation between logarithmic capacity and a modified form of extremal distance.

Let z_0 be a fixed point in the plane, and C_R a circle with center z_0 and radius R, containing E in its interior. Denoting by $\lambda(C_R, E)$ the extremal distance between C_R and E with respect to the interior of C_R, we define

$$\lambda^*(E) = \lim_{R\to\infty} \{\lambda(C_R, E) - 1/2\pi \log R\}. \tag{37}$$

We need not stop here to verify that this limit exists and is finite and independent of z_0, since this will appear from the argument that follows.

Consider the function [12]

$$\rho_0(z) = |\operatorname{grad} g(z)|. \tag{38}$$

For any curve γ joining E and C_R we have

$$\int_\gamma \rho_0 \, |dz| \geq \int_\gamma \partial g/\partial s \, |dz| = \log R + c + \epsilon(R) \tag{39}$$

(independently of z_0, since $\log |z| = \log |z - z_0| + \epsilon(|z - z_0|)$). Further, if $D = (|z - z_0| < R) \cap C(E)$,

$$\begin{aligned} \iint_D \rho_0^2 \, dxdy &= \int_{C_R} g \, \partial g/\partial n \, |dz| \\ &= \int_0^{2\pi} \{\log R + c + \epsilon(R)\}\{1/R + \epsilon(R)\} R \, d\theta \\ &= 2\pi\{\log R + c + \epsilon(R)\}, \end{aligned} \tag{40}$$

again independently of z_0. Hence by the definition of extremal distance, (4) Ch. I,

$$\lambda(C_R, E) \geq 1/2\pi(\log R + c + \epsilon(R)). \tag{41}$$

[12] The following reasoning is similar to that of Chap. I, 3, and we may thus omit details. It follows closely the reasoning for the case of a definition analogous to (37) of a "reduced extremal distance" between E and a finite point, appearing in unpublished work by Ahlfors and Beurling.

Now let ρ be any member of the class P, normalized by $L_\rho(\gamma) \geq 1$. If h denotes the conjugate function of g, we have

$$\int_{h=c} \rho/\rho_0 \, dg = \int_{h=c} \rho \mid dz \mid \geq 1; \tag{42}$$

hence (compare the discussion of the level curves in Ch. I, **3**)

$$\iint_D \rho/\rho_0 \, dgdh \geq 2\pi. \tag{43}$$

Using this and (40) we find (since $\rho_0{}^2$ is the Jacobian)

$$\begin{aligned} 0 &\leq \iint_D (\rho/\rho_0 - 1/(\log R + c))^2 dgdh \\ &\leq \iint_D \rho^2 \, dxdy - 2\pi/(\log R + c) + \epsilon(R), \end{aligned} \tag{44}$$

i. e.

$$\lambda(C_R, E) \leq 1/2\pi \ (\log R + c) + \epsilon(R). \tag{45}$$

Combining (41), (45) and (37) we conclude that

$$\lambda^*(E) = c/2\pi. \tag{46}$$

We now claim that $\lambda^*(E)$ does not decrease if E is circularly symmetrized (see (1)) in any line L and with any point z_0 as center. In view of the definition (36) this can be expressed as follows:

THEOREM.[13] *The logarithmic capacity of a compact set bounded by a finite number of regular curves does not increase under circular symmetrization.*

We shall not take the time to remove the restriction on the boundary of E, but it should be observed that this restriction is used only to simplify the derivation of (46).

For the proof we choose in the theorem (1) E as E_1 and the circle C_R with center at the given point z_0 as E_2, and observe that in this case the extremal distance with respect to the entire plane is the same as that with respect to the interior of E_2. For each R, $\lambda(C_R, E)$ does not decrease in symmetrization, and C_R is its own symmetric image. By the definition (37), our theorem follows.

UNIVERSITY OF KANSAS.

[13] No published proof of this result is known to the author. By comparing the paper by Pólya and Szegö with the indications in the note of Pólya, both referred to in the first footnote of this chapter, one concludes, however, that the result should be familiar to them. Compare also pp. 182-216 in G. Pólya and G. Szegö, "Isoperimetric inequalities in mathematical physics," *Annals of Mathematics Studies*, no. 27, Princeton, 1951, where several results closely connected to the theorems in this chapter are found.

ON GEODESIC TORSIONS AND PARABOLIC AND ASYMPTOTIC CURVES.*

By PHILIP HARTMAN and AUREL WINTNER.

1. *Geodesic torsion.* In the differential geometry of surfaces, the assumption that a surface has a parametrization of class C^2 seems to be a natural one. In fact, it is this assumption that permits the definition of both of the fundamental forms and of the standard curves on the surface (geodesics, asymptotic line, lines of curvature). Assumptions of a higher degree of differentiability for the surface usually have therefore no geometrical significance. Cf. [6], [10]. In the light of this remark, various questions centering about the notion of "geodesic torsion" will be considered in what follows.

Let $X = (x, y, z)$ be a vector in a 3-dimensional Euclidean space and let, in a sufficiently small open domain in a (u^1, u^2)-plane,

$$S:\quad X = X(u^1, u^2) \tag{1}$$

denote a (portion of a) surface of class C^2. By this is meant that $X(u^1, u^2)$ is a function of class C^2 and that the vector product (X_1, X_2), where $X_i = \partial X/\partial u^i$, does not vanish. The unit vector

$$N = N(u^1, u^2) = (X_1, X_2)/|(X_1, X_2)| \tag{2}$$

is the normal vector on (1). The first and second fundamental forms of (1) are defined by

$$ds^2 = dX \cdot dX = g_{ik}du^i du^k \quad\text{and}\quad -dX \cdot dN = h_{ik}du^i du^k, \tag{3}$$

respectively, that is, by

$$g_{ik} = X_i \cdot X_k \text{ and } h_{ik} = -X_i \cdot N_k = X_{ik} \cdot N, \tag{4}$$

where the dots denote scalar multiplications.

Corresponding to any point (u^1, u^2) of S, let $(u^{1\prime}, u^{2\prime})$ represent any pair of numbers for which the vector X' defined by $X_i(u^1, u^2)u^{i\prime}$ is of unit length. Define the vector N' by $N_i(u^1, u^2)u^{i\prime}$, the scalar γ by

$$\gamma = \det(X', N, N'), \qquad (|X'| = 1), \tag{5}$$

and call $\gamma = \gamma(u^1, u^2; u^{1\prime}, u^{2\prime})$ *geodesic torsion.* If a curve Γ: $X = X(s)$

* Received October 25, 1951.

of class C^1 on the surface S passes through the point X of S and has, at that point, the unit tangent vector $X' = dX(s)/ds$, then (5) will be called the geodesic torsion of Γ at the point X.

The geodesic torsion of the point X and the unit vector X' is often defined to be the torsion (at the point X) of a geodesic through X in the direction X'. Two objections can be raised to the latter definition. First, a geodesic on a surface of class C^2 is only of class C^2, while the standard definition of torsion is applicable only to arcs of class C^3. The second objection is that, even if the geodesic is very smooth, its torsion at a point is undefined if the curvature of the geodesic curve vanishes at that point.

The first criticism can be overcome if the word *torsion* of a curve (which need not be on a surface) is defined as in [6], pp. 770-772, where it was shown that torsion can be defined for certain arcs of class C^2 with non-vanishing curvature. In particular, if such an arc has a principal normal of class C^1, then the torsion can be defined geometrically and so as to satisfy the corresponding Frenet equations. As an application of that definition of torsion, suppose that Γ is a geodesic of (1) and has a non-vanishing curvature (at a point, hence near that point). Then it has a principal normal, namely $\pm N$, and the latter is of class C^1 as a function on the arc length on Γ. Since the binormal of Γ is the vector product $(X', \pm N)$, the Frenet equations, as used in [6] (Theorem VI, p. 772), imply the truth of the following assertion:

(I) *On a surface of class C^2, a geodesic arc of non-vanishing curvature possesses a torsion,* (5).

This disposes of the first of the two objections mentioned above. In contrast, the second of those objections cannot be overcome, except by some *ad hoc* definition of what the torsion of a curve should be when the curvature of the latter vanishes. In fact, even on a surface of class C^∞, those points on a geodesic arc, at which the curvature of the latter vanishes, can form a nowhere dense perfect set. Needless to say, a geodesic (or any curve of class C^2 on a surface of class C^2) can have a vanishing curvature at a point only if it is in an asymptotic direction at that point; so that this is the only case excluded in (I).

Whenever the surface (1) is of class C^2, Weingarten's derivation formulae, $N_i = -g^{jk}h_{ij}X_k$, where (g^{ik}) denotes the inverse of the matrix (g_{ik}), are applicable. Hence (5) can be written in the form

$$\gamma = g(g^{2k}h_{ki}u^{i\prime}u^{1\prime} - g^{1k}h_{ki}u^{i\prime}u^{2\prime}), \tag{6}$$

where $g = \det(X_1, X_2, N) = (\det g_{ik})^{\frac{1}{2}} > 0$.

In order to interpret (6), suppose, without loss of generality, the normalizations $g_{11}=g_{22}=1$, $g_{12}=0$ and $h_{12}=0$ at the point (u^1,u^2). Then h_{11} and h_{22} are the principal curvatures, say κ_1 and κ_2, and the directions determined by $(u^{1\prime},u^{2\prime})=(1,0)$ or $0,1)$ are directions of principal curvature, associated with κ_1 or κ_2, respectively. If $(u^{1\prime},u^{2\prime})=(\cos\theta,\sin\theta)$, where θ is the angle from the direction $(1,0)$ of principal curvature, associated with κ_1, to the direction $(u^{1\prime},u^{2\prime})$, then (6) reduces to

$$\gamma=(\kappa_2-\kappa_1)\cos\theta\sin\theta, \tag{7}$$

a formula which is standard (cf. e. g., [3], p. 389) under substantially more severe restrictions than the present assumptions.

It is clear from (7), and from the assumptions under which it was derived above, that the following statement is a corollary:

(II) *If Γ is a curve of class C^1 on a surface of class C^2, then $\gamma=0$ holds at every point of Γ if and only if Γ is a line of curvature.*

2. *On the Beltrami-Enneper theorem.* A point of (1) is called elliptic, hyperbolic or parabolic according as $\det h_{ik}$ (or, what according to (4) is the same thing, the Gaussian curvature $K=\det h_{ik}/\det g_{ik}$) is positive, negative or 0.

If (u^1,u^2) is a non-elliptic point of (1), then a direction $(u^{1\prime},u^{2\prime})$ through the point is called asymptotic if

$$h_{ik}u^{i\prime}u^{k\prime}=0. \tag{8}$$

In the normalization introduced before (7), condition (8) reduces to

$$\kappa_1\cos^2\theta+\kappa_2\sin^2\theta=0. \tag{9}$$

On the other hand, since the product $\kappa_1\kappa_2$ is the Gaussian curvature, (7) is identical with

$$\gamma^2=-K \tag{10}$$

if (9) is satisfied. In view of (7), this proves the following assertion:

(III) *If Γ is an arc of class C^1 on a surface of class C^2, then* (10) *holds at every point of Γ if and only if every tangent vector of Γ either is, or is orthogonal to, an asymptotic direction.*

The theorem of Beltrami-Enneper states that (under certain conditions of smoothness which usually are not, or are erroneously, specified; cf. [6], pp. 773) the relation (10) holds as an identity along asymptotic curves of

non-vanishing curvature. The first assertion of (III) contains this theorem, and (III) avoids the difficulty involved in a statement about the *torsion* (not the *geodesic torsion*) of an asymptotic curve; cf. (IV) below. This difficulty arises when the asymptotic curve is only of class C^1 (or if, when it is smoother, it possesses points at which its curvature vanishes). In this regard, cf. [6], pp. 773-774.

3. *On a result of P. Franklin.* A curve Γ on a surface of class C^2 will be called a parabolic curve if every point of Γ is a parabolic point, that is, if $K = 0$ on Γ.

According to P. Franklin [4], pp. 254-256, a "regular" parabolic curve must be a line of curvature. His definition of a "regular" curve Γ on a surface (1) consists of the following specifications: a) no point of Γ is singular, and b) at no point of Γ does a normal section of the surface S have a flex point; finally c) no point of Γ is a flat point of S (that is, a point at which both factors κ_i of $K = \kappa_1\kappa_2$ vanish).

Franklin's paper was reviewed by Cohn-Vossen and by Rinow. The former [2] points out that Frankin's *proof* is not valid, since it contains a formal error, while the latter [9], without commenting on Franklin's proof, remarks that Franklin's final *assertion* (see above) is not surprising, since the set of the conditions a), b), c) which define "regularity" is quite severe. In what follows, there will be clarified the actual situation resulting from the nature of both of these criticisms. On the one hand, it will be shown that if condition a) is interpreted to mean that grad K (exists and is continuous on S and) does not vanish on the parabolic curve Γ: $K = 0$, then Frankin's *proof* can be saved. On the other hand, it will be shown that condition b) is so severe that the final *assertion* becomes practically vacuous. In fact, the situation proves to be as follows:

(*) *On a surface S of class C^3, let a parabolic curve Γ: $K = 0$ on S be such that* grad $K \neq 0$ *on* Γ, *and suppose that* Γ *satisfies Franklin's condition* b). *Then* Γ *is of class C^2 and a plane curve; as a matter of fact, a plane curve along which the tangent plane of S does not vary; in addition,* Γ *is a line of curvature as well as an asymptotic curve.*

Actually, it cannot even be expected that a parabolic curve will "in general" be a line of curvature. In fact, the notion of a parabolic curve is an intrinsic one, depending only on the metric (g_{ik}), while the notion of a line of curvature depends on the embedding of the metric into the 3-dimensional Euclidean space (that is, on (h_{ik}) as well). The severity of Franklin's con-

ditions is shown by the last italicized statement, a statement which implies that the parabolic curve must become an asymptotic curve (*and* a line of curvature). This situation is the more understandable as the asymptotic curves are the characteristics of the partial differential equation of second order on which the problem of embedding depends.

4. *Asymptotic curves and parabolic curves.* In view of (II) and (III), and of the last italicized statement, (*), which remains to be proved, it seems to be worth while to clarify the relationships between the following three assumptions: (α) Γ *is an asymptotic curve*; (β) Γ *is a line of curvature*; (γ) Γ *is a parabolic curve,* where it is always assumed that Γ is a curve of class C^1 on a surface S of class C^2. It will be shown that (i) *conditions* (α) *and* (β) *imply* (γ) *and that* (ii) *conditions* (α) *and* (γ) *imply* (β), *but that* (iii) *conditions* (β) *and* (γ) *do not imply* (α), *while* (iv) *conditions* (α), (β), (γ) *all are satisfied if and only if* Γ *is a plane curve along which the tangent plane to the surface* S *does not vary.*[1]

The assertions (i) and (ii) are essentially those of Cohn-Vossen [1], pp. 274-275, who, however, involves the extraneous notion of an "envelope" as well as the notion of *torsion,* of an asymptotic curve (and therefore, in particular, heavy restrictions of differentiability).

Proof of (i) *and* (ii). Condition (α) implies, by (III), that $\gamma = \pm(-K)^{\frac{1}{2}}$ on Γ. Hence, $\gamma = 0$ if and only if $K = 0$. It follows therefore from (II) that an asymptotic curve is a line of curvature if and only if it is a parabolic curve.

Proof of (iv). Clearly, condition (α) is equivalent to the assumption of the relation $X' \cdot N' = 0$ along Γ, which means that the tangent vector X' is orthogonal to N' along Γ. On the other hand, the differential equations of Rodrigues for lines of curvatures, $N' + \kappa_i X' = 0$, where κ_i is a principal curvature and $i = 1, 2$, show that X' is parallel to N' along Γ if and only if Γ is a line of curvature. Thus X' is orthogonal to, and at the same time parallel

[1] In his *Introduction to Differential Geometry* (Princeton, 1947), L. P. Eisenhart claims that every plane asymptotic curve is a straight line (p. 249, Ex. 10). That this theorem is false, or that a (smooth) curve Γ satisfying the three conditions (α), (β), (γ) need not be a straight line, is shown by the following example: Consider the curve Γ: $(x, x^2, 0)$ on the surface S: (x, y, z) belonging to $z = (y - x^2)^2$. Along the curve $y = x^2$ of this surface, both z_x and z_y vanish identically. Hence the plane $z = 0$ contains Γ and is the tangent plane to S at every point of Γ. It follows therefore from (iv) [and, of course, from an easy explicit calculation as well] that conditions (α), (β), (γ) are satisfied. But Γ is not a straight line.

to, N' if and only if (α) and (β) hold. However, in this case, since $X' \neq 0$, it follows that $N' \equiv 0$ along the curve Γ. This means that the plane tangent to the surface S is constant along the curve.

Proof of (iii). Conditions (β) and (γ) imply that both $\gamma = 0$ and (10) hold along Γ. Hence (III) shows that Γ is either an asymptotic curve or is orthogonal to an asymptotic direction at every point of Γ. Consequently, (iii) will be proved if it is shown that the second case of this alternative can actually occur. But it is easy to verify that it does occur if the surface (1) is chosen, for instance, as follows: $z = x^2 + y^3$, where $(x, y, z) = X$.

First, the partial derivatives, z_x, z_y and z_{xx}, z_{xy}, z_{yy}, of $z = x^2 + y^3$ are $2x$, $3y^2$ and $2, 0, 6y$, respectively. Hence, from (2)-(4), where $u^1 = x$, $u^2 = v$,

$$(11) \qquad g_{11} = 1 + 4x^2, \qquad g_{12} = 6xy^2, \qquad g_{22} = 1 + 9y^4$$

and, if $\pm(1 + 4x^2 + 9y^4)^{\frac{1}{2}}$ is denoted by j,

$$(12) \qquad h_{11} = 2/j, \qquad h_{12} = 0, \qquad h_{22} = 6y/j.$$

Since $K = \det h_{ik}/\det g_{ik}$, it follows that

$$(13) \qquad K = 12y/(1 + 4x^2 + 9y^4).$$

Consider the curve

$$(14) \qquad \Gamma: \qquad x = x, \qquad y = 0, \qquad z = x^2.$$

This curve is on the surface S: $z = x^2 + y^3$. On the other hand, (12) shows that (8), where $(u^1, u^2) = (x, y)$, reduces to $2dx^2 + 6ydy^2 = 0$, a differential equation for $y = y(x)$ which is not satisfied along the curve (14). Hence (14) is not an asymptotic curve. But it is a parabolic curve, since (13) and (14) imply that $K = 0$. Thus all that remains to be ascertained is that (14) is a line of curvature. This follows, however, by observing that, according to (11) and (12), both g_{12} and h_{12} vanish identically along the coordinate axes, $x = 0$ and $y = 0$, which means that the latter are lines of curvature. Hence the assertion follows from the second of the equations (14).

Condition grad $K \neq 0$ of (*) is satisfied in this example, since the partial derivative of (13) with respect to y is $12/(1 + 4x^2)$, hence distinct from 0, along the curve (14).

5. *Proof of* (*). Consider the surface (1) in a Cartesian parametric form S: $z = (x, y)$, where $z(x, y)$ is a function of class C^3 in a vicinity of $(x, y) = (0, 0)$. It can be supposed that the coordinate axes have been

chosen so that $(x, y) = (0, 0)$ corresponds to a given point of Γ, and that the unit normal vector at this point is directed along the z-axis. Thus

(15) $$z(0,0) = 0 \text{ and } z_x(0,0) = 0, \quad z_y(0,0) = 0.$$

Suppose that the Gaussian curvature, $K = K(x, y)$, vanishes at $(0, 0)$. Then, since

(16) $$K = (z_{xx}z_{yy} - z_{xy}^2)/(1 + z_x^2 + z_y^2)^2$$

for every (x, y), and since $K(0, 0) = 0$, it can be supposed, after a rotation of the (x, y)-plane about the origin, that

(17) $$z_{xy}(0,0) = 0 \text{ and } z_{yy}(0,0) = 0.$$

The last three formula lines show that

(18) $$K_x(0,0) = z_{xx}(0,0)z_{yyx}(0,0) \text{ and } K_y(0,0) = z_{xx}(0,0)z_{yyy}(0,0).$$

Hence, $|\operatorname{grad} K|^2 = K_x^2 + K_y^2$ will not vanish at $(0, 0)$, and therefore at any (x, y) in a vicinity of $(0, 0)$, if and only if

(19) $$z_{xx}(0,0) \neq 0$$

holds and not both $z_{yyx}(0, 0)$ and $z_{yyy}(0, 0)$ vanish. But if assumption b) of (*) is satisfied, then $z_{yyx}(0, 0)$ cannot vanish, since

(20) $$z_{yyy}(0,0) = 0.$$

In fact, if (20) did not hold, then, since (15) and (17) imply that $z(0, y) = z_{yyy}(0, 0)y^3/6 + o(|y|^3)$, it would follow that the normal section $x = 0$ of S: $z = z(x, y)$ has a flex point at $y = 0$.

The equation $K(x, y) = 0$, defining a parabolic curve Γ, and the condition that not both K_x and K_y vanish at a point, say $(0, 0)$, of Γ show that Γ is a curve of class C^1 and satisfies the non-singular differential equation $K_x dx + K_y dy = 0$. According to (18), (19) and (20), this differential equation reduces to $dx/dy = 0$ at $(x, y) = (0, 0)$. But (17) and (19) imply that $dx/dy = 0$ defines the (unique) asymptotic direction at $(0, 0)$. Hence, dx/dy is an asymptotic direction at $(0, 0)$. Since $(0, 0)$ represents an arbitrary point of Γ, it follows that Γ is an asymptotic curve. Hence Γ is of class C^2 and (5) follows from (i), (ii) and (iii).

6. *On a theorem of van Kampen.* Using a fallacious generalization of the Beltrami-Enneper formula, $\tau^2 = K$, for the torsion, τ, of an asymptotic curve, van Kampen [8] has arrived at the following result:

(†) *Let S be a surface of class C^2, and P a hyperbolic point ($K < 0$) of S. Let J denote one branch of the intersection of S and of the plane*

tangent to S at P, and suppose that at every point distinct from P the plane curve J has a non-vanishing curvature. Then there exists on S at least one asymptotic curve, say Γ, which passes through P in such a way that J is tangent to Γ at P and lies between Γ and the common tangent; that is, J lies between Γ and the normal section of S determined by the common direction of J and Γ at P.

That van Kampen's generalization of the Beltrami-Enneper formula is false, is seen by comparing it (Theorem (I) in [8]) with Bonnet's formula (cf. (44) below), which connects the curvature and the torsion of a curve drawn through a point of a surface in an asymptotic direction. The error is made when, in the last sentence of the second paragraph of [8], p. 992, van Kampen assumes that he can differentiate a relation (*along* a curve), whereas the relation in question is valid only at *one* point of that curve.

It turns out, however, that van Kampen's final result, represented by the last italicized statement, (†), happens to be correct. Tis will be proved in what follows.

Remark. It will remain undecided whether, in the above wording of (†), the passage "*at least one* asymptotic curve" can be replaced by "*all* asymptotic curves" or, for that matter, by "*the* asymptotic curve." In this regard, cf. the example, given in [7], pp. 153-156, of a surface S of class C^2 having a negative curvature K and containing a point P which issues more than one asymptotic curve (hence a continuum of asymptotic curves) in the same asymptotic direction. In that particular example, there exist asymptotic curves on different sides of their (common) tangent at P but, in contrast to the assumption in (†) above, the curvature of the curve of intersection, J, vanishes at a sequence of points which cluster at P. Hence it remains a question whether or not the non-vanishing of the curvature of J (at those points of J which are distinct from P) implies that all asymptotic curves (touching J at P) are separated by J from the tangent of J at P or, for that matter, that the asymptotic curve (through P and in the direction of J) must be unique.

It may be mentioned that (†) could also be deduced from Beltrami's formula (cf. (50) below), *if* S is of class C^3 and J has a non-vanishing curvature.

7. *Proof of* (†). It can be assumed that S is given in the form $z = z(x, y)$, where $z(x, y)$ is a function of class C^2; that P is at $(x, y) = (0, 0)$ and that (15) is satisfied; finally that, since the value of (16) at P is

supposed to be negative, $z_{xx}(0,0) = z_{yy}(0,0) = 0$, while $z_{xy}(0,0)$ is positive. If (by the choice of the units of length along coordinate axes) this positive number is chosen to be 1, it follows that

$$z = z(x,y) = xy + o(x^2 + y^2), \tag{21}$$

where the o-term represents a function of class C^2 in (x,y). The form on the left of (8) reduces at P to $x'y'$, a form which represents positive values for directions corresponding to the first and third quadrants, $(x' > 0, y' > 0)$ and $(x' < 0, y' < 0)$, and negative values for the second and fourth quadrants. In particular, the asymptotic directions through P are the directions of the coordinate axes.

Consider that branch, J, of the intersection of S and of the tangent plane at P (i.e., of the plane $z = 0$) which is tangent to the x-axis. Then, by Lemma (§) of Section 8 below, J is a curve which (for small $|x|$) can be represented in the form

$$J\colon\ y = y(x), \tag{22}$$

where $y(x)$ is a function of class C^1 for all x, and of class C^2 for all $x \neq 0$; cf. the assumption of (†) concerning the non-vanishing of the curvature of J at points distinct from P.

According to (21), and since (22) is a curve passing through $P = (0,0)$, the partial derivative $z_y(x,y)$ for small positive x is $x + o(x)$ along the curve J. Hence, if x in (22) is chosen to be positive, then $z_y(x, y(x))$ is positive (for small $x > 0$). On the other hand, since the curve (22) is on the surface $z = z(x,y)$ and on the plane $z = 0$,

$$z(x, y(x)) = 0, \tag{23}$$

and differentiation of (23) with respect to x gives

$$z_x(x, y(x)) + z_y(x, y(x))y'(x) = 0 \tag{24}$$

(even if $x = 0$), whence one more differentiation leads (if $x \neq 0$) to

$$z_{xx} + 2z_{xy}y' + z_{yy}y'^2 + z_y y'' = 0, \qquad \text{where } y = y(x). \tag{25}$$

Since the coeffiicient, z_y, of y'' in (25) was just seen to be positive (for small positive x), it follows that the sum of the first three terms of (25) is positive or negative according as $y''(x)$ is negative or positive for small positive x (in fact, $y''(x) = 0$ is excluded by the assumption of a non-vanishing curvature at the points of (22) distinct from its point belonging to $x = 0$).

Suppose, for instance, that $y''(x) > 0$ (for small positive x). Then $[y(x)] < 0$, if $[y(x)]$ denotes the sum of the first three terms of (25).

On the other hand, (8) shows that if $y = y^*(x)$, $z = z(x, y^*(x))$ is an asymptotic curve, then $[y^*(x)] = 0$. It follows therefore from the signature rule of the four quadrants (described, after (21), with regard to the second fundamental form of the surface at the point P), that, for small positive x, the slope of the relevant asymptotic direction at the point $(x, y(x))$ of the surface is greater than the slope of the curve (21) at the corresponding x. In fact, this is clear for reasons of continuity, since the asymptotic directions at the point $(x, y(x))$ of the surface are close to those at the point $P = (0, 0)$, while the latter directions are those of the coordinate axes, $x = 0$ and $y = 0$.

Accordingly, if α is any point of an interval $0 < \alpha < a$, where $a > 0$ is sufficiently small, then there is an asymptotic curve (at least one), say

(26) $$\Gamma_\alpha\colon\ y = y(x; \alpha),$$

passing through the point $(\alpha, y(\alpha))$ of the curve (22) in a direction nearly tangent to the direction of the curve (22) at $x = \alpha$, and all functions (26) of x will exist for $0 \leqq x \leqq a$ (if $0 < \alpha < a$, and if a is sufficiently small), finally all these curves (26) will satisfy, with reference to the curve (22), the inequality

(27) $$y(x; \alpha) > y(x) \text{ if } \alpha < x \leqq a.$$

Since the surface S is hyperbolic at (hence near) the point $P = (0, 0)$, its asymptotic curves (26) will satisfy, for $0 \leqq x \leqq a$, a differential equation (8), which is of the form

(28) $$y' = f(x, y),$$

where $f(x, y)$ is continuous (for small $x^2 + y^2$). Hence the standard arguments, which deal with (28) on the basis of equicontinuous functions, show that there exists a sequence of positive numbers $\alpha_1, \alpha_2, \cdots$ satisfying $a > \alpha_n \to 0$, as $n \to \infty$, and having the property that $y(x; \alpha_n)$, the function (27) belonging to $\alpha = \alpha_n$, tends to a limit function, uniformly for $0 \leqq x \leqq a$, as $n \to \infty$. Furthermore, this limit function $y = y^*(x)$ is (as is each of the curves (26)) the graph of the projection on the (x, y)-plane of an asymptotic curve. Finally, it follows from (27) that

(29) $$y^*(x) \geqq y(x) \qquad \text{for } 0 \leqq x \leqq a.$$

Clearly, (29) completes the proof of the last italicized statement, (†), if the lemma (§) of Section 8 is granted. In fact, while (29) was deduced for the case in which $y''(x) > 0$ for small $x > 0$, the case in which $y''(x) < 0$ for small $x > 0$ (as well as both cases of a small $-x > 0$) can, of course, be treated in the same way.

8. *On the "true" Dupin diagram.* Let T be a plane through a hyperbolic point P of a surface S of class C^2. If T is not the plane tangent to S at P, and if S is a sufficiently small neighborhood of P, then the set ST, along which T intersects S, consists of a Jordan arc of class C^2. This is an immediate consequence of (the C^2-form of) the classical theorem on implicit functions, the exemption of the tangent plane being equivalent to the non-vanishing of the gradient involved. Correspondingly, if T is the tangent plane, then, since the gradient vanishes at P, no general theorem on implicit functions is applicable to either of the branches of which, in view of Dupin's indicatrix, the set ST can be expected to consist. That the essential (but, perhaps against expectation, not all) aspects of what is indicated by Dupin's approximation is nevertheless true, is the content of the first two assertions of the following lemma (the third assertion of which shows that, because of the vanishing of the gradient, the C^2-character of the "implicit functions" can actually be lost).

(§) *If S is a sufficiently small neighborhood of a hyperbolic point P on a surface of class C^2, and if T denotes the plane tangent to S at P, then*

(i) *the intersection ST consists of two Jordan arcs, say J_1 and J_2, each of which contains P in its interior, and P is the only point common to J_1 and J_2;*

(ii) *both plane curves J_1, J_2 have continuous tangents (also at P) and, except possibly at P, continuous curvatures as well;*

(iii) *the curvature of J_i at P need not exist (and, if it exists at P, it need not be continuous at P)*, unless an assumption going beyond the C^2-assumption (such as the C^3-assumption) is required of S.

Assertion (i) and (at least the first part of) assertion (ii) are closely connected with the results of Hadamard [5] on the invariant curves of a surface transformation near a fixed point of hyperbolic type. It is, however, more convenient to prove (i) and (ii) directly, and in a way which, in contrast to Hadamard's procedure, does not depend on the method of successive approximations.

As at the beginning of Section 7, it can be assumed that S is given in the form $z = z(x, y)$, and that T is the (x, y)-plane and P is its point $(0, 0)$, but that (21) is replaced by

$$S\colon z = z(x, y) = \tfrac{1}{2}(y^2 - x^2) + o(x^2 + y^2), \tag{30}$$

where the o-term represents a function which is of class C^2 in (x, y) (in fact, (30) differs from (21) in a rotation about P). According to (30), the

asymptotic directions of S at P, represented by the asymptotes of Dupin's hyperbola at P, are the bisectors ($x \pm y = 0$) of the coordinate quadrants.

Since T is the plane $z = 0$, its intersection with S is the set of points (x, y) satisfying

$$(31) \qquad ST: \tfrac{1}{2}(y^2 - x^2) + o(x^2 + y^2) = 0.$$

It is clear from (31) that, if S, or the (x, y)-domain under consideration, is chosen small enough, then every point of ST is contained in one of the four wedges (issuing from $(0, 0)$ and bisected by the four asymptotic half-lines) which are defined by the inequalities $\tfrac{1}{2}|x| \leqq |y| \leqq 2|x|$.

Consider the wedge contained in the first quadrant of the (x, y)-plane, that is, the wedge

$$(32) \qquad \tfrac{1}{2}x \leqq y \leqq 2x \qquad \text{(so that } x > 0 \text{ and } y > 0 \text{ unless } x = 0 = y\text{)}.$$

It will be shown that those points (x, y) of ST which are contained in this wedge form (for sufficiently small $x \geqq 0$) a Jordan arc which is representable in the form $y = y(x)$, where $y(x)$ is a (single-valued) function having a continuous first derivative $y'(x)$; that $y'(0)$ (when interpreted as the derivative at $x = 0$ from the right) is 1; finally that the function $y(x)$ has a continuous second derivative $y''(x)$ if $x > 0$. Since (32) could be replaced by any of the four wedges, this will prove assertions (i) and (ii) of the last italicized assertion.

Proof of (i)-(ii). It is clear from (30) that, if x is positive (and small enough), then $z(x, \tfrac{1}{2}x)$ and $z(x, 2x)$ are of opposite sign. Hence $z(x, y)$ must vanish at least once, say at the ordinate $y = y(x)$, when $x > 0$ is fixed and y varies from the lower to the upper half-line bordering the wedge (32). On the other hand, since the o-term in (30) is a function of class C^2 in (x, y), it is clear from (30) that $z_y(x, y) = y + o(x^2 + y^2)^{\frac{1}{2}}$. Hence, if $x > 0$ is small enough, $z_y(x, y)$ is positive within the wedge. Consequently, the ordinate $y = y(x)$, mentioned before, is unique. It now follows by standard arguments, occurring in the proof of the classical theorems on implicit functions, that the function $y(x)$ is continuous for $x \geqq 0$ and that it has a continuous second derivative for $x > 0$; cf. (24) and (25). That it has a continuous first derivative at $x = 0$ also, follows from (24). In fact, (30) shows that (24) can be written in the form

$$-x + o(x) + (y(x) + o(x))y'(x) = 0$$

for small positive x, and (31) shows that $y(x)/x = 1 + o(1)$.

Proof of (iii). In a neighborhood of $x = 0$, let $f(x)$ be any function

satisfying the following pair of conditions: (a) there exists a continuous second derivative $f''(x)$ (also at $x=0$); (b) if $x \to \pm 0$, then $f(x)=o(x^2)$. In terms of such an $f(x)$, define S by $z=z(x,y)$, where $z(x,y)=xy+f(x)$. Then S is of class C^2, by (a), and (21) is satisfied, by (b). Furthermore, it is seen that (23), the equation defining the intersection ST, splits into J_1: $x=0$ and J_2: $y=f(x)/x$. Hence, in order to conclude the truth of (iii), it is sufficient to observe that there exist two functions, say $f(x)=g(x)$ and $f(x)=h(x)$, which satisfy both (a) and (b) and have the property that the function defined by $y(x)=f(x)/x$ (if $x\neq 0$, and by $y(0)=0$ if $x=0$) has at $x=0$ no second derivative or a discontinuous second derivative according as $f=g$ or $f=h$.

9. *Geodesic curvature and geodesic torsion.* After the italicized assertion, (†), of Section 6, reference was made to a formula due to Bonnet (cf. [3], pp. 397-399). Inasmuch as this formula, which is (44) below, is usually derived in a somewhat roundabout way and without a specification of the assumptions on which it depends, it will be proved in what follows by a more direct approach, leading to a reasonable minimum of the conditions to be required for its validity.

Let S be a surface of class C^2, and Γ: $X=X(s)$ a curve of class C^2 on Γ, where s denotes the arc length on Γ. Define on Γ three, mutually perpendicular, unit vectors V_1, V_2, V_3 (which form a right-hand orthogonal system), by placing

$$(33) \qquad V_1(s)=X'(s), \qquad V_2(s)=N(s), \qquad V_3(s)=(X'(s),N(s)),$$

where the prime denotes differentiation with respect to s and N is the surface normal, (2), expressed along Γ as a function of s. Clearly, all three functions $V_i(s)$ are of class C^1. Hence the three derived vectors, V_i', exist, are continuous, and are linear combinations, with continuous scalar coefficients, of the three vectors V_k. These coefficients can be calculated as is usual for all "derivation formulae." This leads to the well-known "geodesic Frenet equations,"

$$(34) \qquad V_1'=\alpha V_2-\beta V_3, \qquad V_2'=-\alpha V_1+\gamma V_2, \qquad V_3'=\beta V_1-\gamma V_2,$$

where

$$(35) \qquad \alpha=X''\cdot N, \qquad \beta=\det(X',X'',N), \qquad \gamma=\det(X',N,N').$$

The latter γ is identical with the γ in (5), which in Section 1 was defined to be the *geodesic torsion.* Correspondingly, the second of the relations (35) shows that β is identical with the classical ("embedded") definition of the

geodesic curvature. Finally, it is seen from (4) that the first of the relations (35) can be written in the form

$$(36) \qquad \alpha = -X' \cdot N'; \text{ hence } \alpha = h_{ik}u^{i'}u^{k'},$$

by (3) (so that α is the *normal curvature*). Thus $\alpha = 0$ is equivalent to (8), which is the definition of an asymptotic direction (if any; that is, if $K \leqq 0$).

If points of Γ at which $|X''|$ may vanish are excluded, then the set (33) (in which the assumption $|X''| > 0$ is not needed) can be paralleled by the set consisting of the unit vectors of the tangent, principal normal, and binormal, of Γ, that is, by the set

$$(37) \qquad U_1 = X', \qquad U_2 = |X''|^{-1}X'', \qquad U_3 = |X''|^{-1}(X', X'').$$

Then the "geodesic" Frenet equations, (34), become replaced by the ordinary Frenet equations,

$$(38) \qquad U_1' = \kappa U_2, \qquad U_2' = -\kappa U_1 + \tau U_3, \qquad U_3' = -\tau U_2,$$

and, correspondingly, the data (35) by the (ordinary) *curvature* and the (ordinary) *torsion*,

$$(39) \qquad \kappa = |X''| > 0 \text{ and } \tau = \det(X', X'', X''')/\kappa^2,$$

provided that Γ: $X = X(s)$ (instead of being, as before, just of class C^2) is of class C^3. But the definitions (37) and the first of the relations (38), where $\kappa = |X''| > 0$, do not require the latter proviso, and imply, in view of (37), the identities

$$(40) \quad U_1 = V_1, \quad U_2 = V_2 \cos\omega + V_3 \sin\omega, \quad U_3 = -V_2 \sin\omega + V_3 \cos\omega,$$

if the (continuous) angular function $\omega = \omega(s)$ is defined (mod 2π) by

$$(41) \qquad \cos\omega = N \cdot U_2, \qquad \sin\omega = -\det(U_1, U_2, N).$$

If this is compared with (37) and (38), it follows that

$$(42) \qquad \alpha = \kappa\cos\omega, \qquad \beta = -\kappa\sin\omega.$$

Since the definitions of α, β and κ imply that $\alpha = 0 = \beta$ if $\kappa = 0$, both equations (42) hold for $\kappa = 0$ also, provided that the angle ω, which the case $\kappa = 0$ of (41) leaves undefined, is considered as arbitrary.

It will now be supposed that $\kappa > 0$, and that Γ is of class C^3. Then (38) is applicable and, in view of (41), the continuous function $\omega = \omega(s)$ is of class C^1, as is $\kappa = \kappa(s)$. Hence, if (40) is differentiated, and if the result is compared with (34) and (38), it follows that

$$(43) \qquad \omega' = \tau - \gamma.$$

A corollary of (36), (42) and (43) is the following pair of facts:

(IV) *Let S be a surface, of class C^3, having no elliptic points (so that $K \leqq 0$ on S), and let Γ be a curve, of class C^2, on S. Then Γ is an asymptotic curve if and only if $\kappa(s) \equiv |\beta(s)|$, where $\kappa(\geqq 0)$ is the curvature, and β the geodesic curvature, of Γ. If, in addition, $\kappa(s) \neq 0$ holds on an asymptotic curve Γ, then $\tau(s) \equiv \gamma(s)$, where $\tau(s)$ is the torsion, and $\gamma(s)$ the geodesic torsion, of Γ.*

Since S is supposed to be of class C^3, the asymptotic curves are of class C^2. Hence, the classical definition, (39), of τ will not in general apply (cf. [6], pp. 773). But if the torsion, τ, is defined as in [6], pp. 770-772, then the asymptotic curve has a torsion at all those points s at which $\kappa(s) \neq 0$. This, and only this, makes meaningful the second assertion of (IV).

Of course, the converse of the second assertion of (IV) is false, that is, the assumption $\tau(s) \equiv \gamma(s)$ along a curve of class C^2 (with non-vanishing curvature) does not imply that the curve is an asymptotic curve. In fact, this identity holds, under the assumption $\kappa(s) > 0$, if and only if $U_2(s) \cdot N(s) =$ const. on Γ (a condition which is satisfied if, for instance, Γ is an asymptotic curve or a geodesic).

Let s be fixed. Then, if $\kappa(s) = |X''(s)|$ vanishes, (35) shows that $\alpha(s) = \beta(s) = 0$. On the other hand, if $\kappa(s) > 0$, then the first of the relations (42) shows that $\alpha(s) \equiv 0$ is equivalent to $\cos \omega(s) \equiv 0$, which implies that $\omega(s)$ is a multiple of π, and that $\omega'(s)$ therefore exists and is 0 (in a vicinity of the fixed s). In other words, $\alpha(s) \equiv 0$ implies that $\kappa \equiv |\beta|$, by the second part of (42), and that $\tau \equiv \gamma$ if $\kappa \neq 0$, by (43); conversely, $\kappa \equiv |\beta|$ implies that $\alpha \equiv 0$. This completes the proof of (IV), since, in view of the remark made after (36), the differential equation of the asymptotic curves is $\alpha = 0$.

10. *On a formula of Bonnet.* It will now be easy to formulate a precise wording of Bonnet's relation (cf. [3], pp. 397-399), referred to at the beginning of Section 9.

(V) *Let S be a surface of class C^3 having no elliptic points (so that $K \leqq 0$), and let Γ be a curve on S which is of class C^3 and has, at some point P, an asymptotic direction and a non-vanishing curvature. Then*

$$(3\tau_0 - \tau)\kappa = \pm 2\tau_0\kappa_0, \qquad (\kappa > 0, \kappa_0 \geqq 0), \tag{44}$$

where κ and τ denote curvature and the torsion, and $\pm \kappa_0$ and $\tau = \pm(-K)^{\frac{1}{2}}$ the geodesic curvature and the geodesic torsion, of Γ at P (so that, if Γ_0 denotes

the asymptotic curve tangent to Γ at P, then κ_0 is the curvature and, if $\kappa_0 > 0$, then τ_0 is the torsion of Γ_0 at P; cf. (IV) above).

When $\kappa\kappa_0 > 0$, the alternative sign $(\pm)$ in (44) depends on Γ; in fact, it will be clear from the proof of (44) that the $+$ or the $-$ holds according as the principal normals of Γ and Γ_0 have common or opposite directions.

If P is a parabolic point of S, then any curve Γ (of class C^3) through P having an asymptotic direction possesses, at P, either a vanishing curvature κ or a vanishing torsion τ, since $\kappa > 0$ and $\tau_0 = \pm(-K)^{\frac{1}{2}} = 0$ imply that $\tau = 0$ in (44).

The proof of the last italicized statement, (V), proceeds as follows:

First, a multiplication of the two determinants in (35) shows that $\beta\gamma$ can be written as the determinant in which the first row is 1, 0, $X' \cdot N'$, the second $0, X'' \cdot N, X'' \cdot N'$, and the third 0, 1, 0. Hence $\beta\gamma = -X'' \cdot N'$. Since differentiation of the first of the relations (36) gives $\alpha' = -X'' \cdot N' - X' \cdot N''$, it follows that

$$\alpha' - 2\beta\gamma \tag{45}$$

is identical with $X'' \cdot N' - X' \cdot N''$.

Since the curve Γ: $X = X(s) = X(u^1(s), u^2(s))$ is on the surface S: $X = X(u^1, u^2)$, differentiations of (1) and (2) with respect to s show that $Z' = Z_i u^{i\prime}$ and $Z'' = Z_{ik} u^{i\prime} u^{k\prime} + Z_i u^{i\prime\prime}$ hold for $Z = X$ and for $Z = N$ (the subscripts denote partial differentiation with respect to u^i, u^k). Hence, the expression (45) is the sum of

$$(X_{ik} \cdot N_j - X_j \cdot N_{ik}) u^{i\prime} u^{k\prime} u^{j\prime} \tag{46}$$

and of the bilinear form $h_{ik}(u^{i\prime\prime} u^{k\prime} - u^{k\prime\prime} u^{i\prime})$, where h_{ik} is the scalar product defined by (3) or (4). This bilinear form vanishes identically, since $h_{ik} = h_{ki}$. Consequently, the function (45) or (46) of s depends only on the point P: $X(s)$ and on the direction, $X'(s)$, of Γ at P.

The identity of the two values (45), (46) was just derived under the hypothesis that Γ is of class C^2. If Γ is of class C^3, so that $\kappa(s)$ and $\omega(s)$ are of class C^1, then (42) shows that (46) can be written in the form

$$\kappa' \cos\omega - (\omega' - 2\gamma)\kappa \sin\omega. \tag{47}$$

Let this be applied to both Γ_0 and Γ, where Γ_0 is an asymptotic curve through a point, P, of S, and Γ is a curve, of class C^3 and of non-vanishing curvature, which is on S and is tangent to Γ_0 at P. In view of the identity of the numbers (45), (46), it is seen that (45) attains, at P, the same

value for Γ as for Γ_0. Hence, if $\kappa, \omega, \gamma, \cdots$ and $\kappa_0, \omega_0, \gamma_0, \cdots$ refer to Γ and Γ_0, respectively, then the value of the expression (45) at the point P is

$$(48) \qquad -2\beta\gamma = 2\kappa_0\tau_0 \sin\omega_0, \qquad (\sin\omega_0 = \pm 1),$$

since $\alpha(s) \equiv 0$, hence $\alpha'(s) \equiv 0$, on Γ_0.

On the other hand, since Γ is of class C^3, another expression for the value of (45) at P is given by (47). In view of (43), and since $\gamma = \tau_0$, the identity of the values (47), (48) means that

$$(49) \qquad \kappa' \cos\omega + (3\tau_0 - \tau)\kappa \sin\omega = 2\kappa_0\tau_0 \sin\omega_0.$$

Finally, since Γ has an asymptotic direction at P, it follows that $\alpha(s) = 0$ at P. Hence, the first relation of (42) and the assumption $\kappa(s) \neq 0$ show that $\cos\omega = 0$, $\sin\omega = \pm 1$. Consequently, (44) follows from (49).

11. *On a formula of Beltrami.* Beltrami's theorem, mentioned at the end of Section 6, states that if Γ is a branch of the intersection, ST, of S and of the plane, T, tangent to S at a hyperbolic point, P, of S, then, in the notations of (V) above,

$$(50) \qquad 3\kappa = 2\kappa_0$$

(cf. [3], p. 398). Formally, this result of Beltrami is a consequence of Bonnet's theorem, since (44) reduces to (50) if $\tau = 0$ (in fact, since $\tau_0 = \pm(-K)^{\frac{1}{2}}$, hence

$$(51) \qquad \tau_0 \neq 0$$

at a hyperbolic point, division by τ_0 is allowed).

Actually, this deduction of (50) from (44) is not legitimate under the assumptions of (V). For, on the one hand, (50) is claimed also for the case $\kappa = 0$, excluded in (38), and, on the other hand, (V) assumes that Γ is of class C^3, whereas, corresponding to (iii) in (§), Section 8, the curve Γ need not be of class C^3 under the C^3-assumption made in (V) for S. It will, however, be shown that the proof of (V) can be adjusted to the case of Beltrami's theorem so as to dispose of the difficulties on both of these accounts; so that (50) holds in its full generality:

(VI) *Let P be a hyperbolic point of a surface S of class C^3, and let Γ denote a branch of the intersection ST, where T is the plane tangent to S at P. Then Γ has at P a (continuous) curvature, $\kappa \geqq 0$, and (50) holds for the curvature, $\kappa_0 \geqq 0$, of the asymptotic curve tangent to Γ at P.*

In order to prove this assertion, (VI), suppose first that $\kappa(s)$, the curvature of Γ, vanishes on a set of points which cluster at P. It then follows from (42) that, for reasons of continuity, $\kappa = \beta = 0$ at P. On the other hand, it was shown in Section 10 that α' exists even under the present assumptions, and so it is seen from (42) that α' vanishes at P. Consequently, the same is true of the expression (45), and therefore of the expression (46) as well, and so of (48). But the vanishing of (48) at P is equivalent to $\kappa_0 = 0$, since (51) is satisfied (cf. the corresponding conclusion in Section 10). Hence (50) is true in the present case.

In the remaining case, $\kappa(s)$ does not vanish near P (it may or may not vanish *at* P), hence (41) defines $\omega(s)$ near P. It follows therefore from an obvious variant of the lemma, (§), of Section 8 that Γ is of class C^3 near P. But $\tau(s) \equiv 0$, since $\Gamma = ST$ is a plane curve. Thus it is clear from (43) that (whether $\kappa(s)$ does or does not vanish at P) it is possible to define ω (*at* P) in such a way that $\omega(s)$ remains of class C^1 after the inclusion of P; in fact,

$$\omega' = -\gamma = -\tau_0 \text{ at } P. \tag{52}$$

On the other hand, since Γ is in the plane tangent to S at P, it is clear that, *near* P, the vector product (U_1, U_2) has the constant direction of the normal to S at P (note that the principal normal, U_2, is undefined *at* P if $\kappa = 0$). It follows therefore from (41) that

$$\sin\omega = \pm 1 \text{ and } \cos\omega = 0 \text{ at } P, \tag{53}$$

and it is clear from (53) that $\alpha(s) = \kappa(s)\cos\omega(s)$ is differentiable at P, having the derivative

$$\alpha' = 0 + \kappa(\cos\omega(s))'_P = -\kappa\omega'\sin\omega \text{ at } P \tag{54}$$

(even though $\kappa(s)$ may not be differentiable at P). From now on, all statements will refer to the point P.

Since (54), (52) and the second of the relations (42) imply that $\alpha' = \kappa\tau_0\sin\omega = -\beta\gamma$, the value of the expression (45) is $3\kappa\tau_0\sin\omega$. On the other hand, the expression (45) is identical with (46) as well as with (48). Consequently, the product on the right of (48) must have the value $3\kappa\tau_0\sin\omega$, just found. In view of (51), this means that $2\kappa_0\sin\omega_0 = 3\kappa\sin\omega$, which, according to (53) and the parenthetical alternative in (48), implies that $|2\kappa_0| = |3\kappa|$. This proves (50), since both curvatures κ, κ_0 are non-negative in (VI).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] S. Cohn-Vossen, "Die parabolische Kurve," *Mathematische Annalen*, vol. 99 (1928), pp. 273-308.

[2] ———, *Zentralblatt für Mathematik*, vol. 9 (1934), pp. 373-374.

[3] G. Darboux, *Leçons sur la théorie générale des surfaces*, vol. 2 (1889).

[4] P. Franklin, "Regions of positive and negative curvature on closed surfaces," *Journal of Mathematics and Physics*, vol. 13 (1934), pp. 253-260.

[5] J. Hadamard, "Sur l'itération et les solutions asymptotiques des équations différentielles," *Bulletin de la Société Mathématique de France*, vol. 29 (1901), pp. 224-228.

[6] P. Hartman and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[7] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, vol. 73 (1951), pp. 149-172.

[8] E. R. van Kampen, "A remark on asymptotic curves," *ibid.*, vol. 61 (1939), pp. 992-994.

[9] W. Rinow, *Jahrbuch über die Fortschritte der Mathematik*, vol. 60_1 (1934), pp. 613-614.

[10] A. Wintner, "On isometric surfaces," *American Journal of Mathematics*, vol. 74 (1952), pp. 198-214.

ON THE THEORY OF GEODESIC FIELDS.*

By Philip Hartman and Aurel Wintner.

1. *Geodesics.* On a sufficiently small (u, v)-domain, say on

$$(1) \qquad \mathcal{B}_a\colon \quad u^2 + v^2 < a^2,$$

consider the Riemannian geometry defined by a line element

$$(2) \qquad ds^2 = E(u, v)\,du^2 + 2F(u, v)\,du\,dv + G(u, v)\,dv^2,$$

which is positive definite (i. e.,

$$(3) \qquad EG - F^2 > 0 \text{ and } E > 0,$$

hence $G > 0$), but such that the functions E, F, G of (u, v) are just continuous. A geodesic Γ must then be defined as a Jordan arc contained in $\mathcal{B}_a$ and having the property that, if P and Q are the end points of Γ, and if Λ is any Jordan arc joining P to Q within $\mathcal{B}_a$, then

$$\int_\Gamma |\,ds\,| \leqq \int_\Lambda |\,ds\,|.$$

If P and $Q \neq P$ are close enough to the center, $(0, 0)$, of the (u, v)-domain (1), then, according to Hilbert ([9]; cf. [2], pp. 419-438), there exists a geodesic $\Gamma = \Gamma(P, Q)$, which is a rectifiable Jordan arc. (Actually, Hilbert assumes that the line element (2) is embedded into a Euclidean (x, y, z)-space as the $dx^2 + dy^2 + dz^2$ on a surface of class C^1, but this assumption is nowhere used in his proof.)

One of the difficulties is that, no matter how close Q be chosen to a fixed P, the geodesic joining Q to P need not be unique, not even if the coefficient functions of (3), instead of being just continuous, are of class C^1; cf. [7], pp. 132-133. If they are of class C^1, then the Christoffel coefficients $\Gamma^i_{jk} = \Gamma^i_{jk}(u, v)$ exist and are continuous, and, as pointed out in [7], pp. 134-135, all geodesics must satisfy the standard differential equations $u^{i\prime\prime} + \Gamma^i_{jk} u^{j\prime} u^{k\prime} = 0$, where $(u^1, u^2) = (u, v)$.

Since the coefficient functions of (2) will not be required to possess derivatives, no differential equations for the geodesics will be available. Nevertheless, there will be proved several central theorems of the classical theory

* Received November 12, 1951.

(a theory the methods of which assume the coefficient functions of (2) to be of class C^2, at least, and often even smoother). The theorems in question are those of Gauss on orthogonal trajectories, the related theorem of Jacobi concerning multipliers, Riemann's invariant relations which are both necessary and sufficient (Herglotz) for normal geodesic coordinates, and Beltrami's characterization of the non-Euclidean and spherical metrics as geodesic maps of the Euclidean geometry.

Even if the coefficient functions of (2) were assumed to be of class C^1, the metric defined by (2) would not, in general, have a Gaussian curvature $K = K(u, v)$ (in fact, the classical definition of K applies only if E, F, G are of class C^2). In particular, Jacobi's equation of the normal displacements, $d^2n/ds^2 + K(s)n = 0$, which defines conjugate points, is not available.

In problems dealing with the geodesics of a metric (2) in which E, F, G are just continuous, a complication is presented by the circumstance that a geodesic arc Γ cannot be assumed to be a curve of class C^1 or, for that matter, such as to possess a tangent at each of its points (rather than just almost everywhere, Γ being rectifiable). Actually, these questions were left undecided in [7], pp. 144-148. (It was shown there, among other things, that every Γ must possess the "Archimedian property"; and while it is easy to see that this property of a rectifiable Jordan arc does not imply the existence of a tangent at every point of the arc,[1] it seems to be less easy to "embed" such an arc into a metric (2) for which the arc becomes a geodesic.) Fortunately, these matters will not complicate the situation, since the nature of all the problems considered is such as to imply the C^1-character of those particular (even though possibly not of all) geodesics Γ of a metric (2) to which the assertions of the theorems refer.

2. *Transversals.* The principal results will depend on (the wording or the proof of) a lemma concerning transversal fields; cf. [6], pp. 147-151. Under the classical assumptions of differentiability, the assertion of the lemma is nothing but a theorem of Jacobi, and is then a corollary of the theorem of Gauss concerning the transversal trajectories of a sheaf of geodesics (cf.

---

[1] In order to obtain such a Jordan arc in a (u, v)-plane, it is sufficient to put $u = x \cos \phi$, $v = r \sin \phi$, and to assign the Jordan arc in the parametric form $r = f(\phi)$, where $f(\phi)$, with $f(\phi_1) \neq f(\phi_2)$ when $\phi_1 \neq \phi_2$, is positive, is of class C^1 for $-\infty < \phi < \infty$ and tends sufficiently fast to 0 as $\phi \to \pm \infty$ (so that the point $(u, v) = (0, 0)$ of the Jordan arc belongs to $\phi = -\infty$ *and* $\phi = \infty$). In fact, an easy calculation shows that the choice $f(\phi) = \exp(-\phi^2)$ will do (the slower logarithmic spiral, $f(\phi) = \exp(-|\phi|)$, will not do; the non-differentiability of the latter function at $\phi = 0$ is of course immaterial).

[3], vol. II, p. 430). But the point is that, under the general assumption (*) below, the classical proofs must fail to apply. The lemma in question is as follows:

LEMMA 1. *Suppose that*

(0) *E, F, G in* (2) *are continuous functions satisfying* (3) *on a sufficiently small domain* (1), *and that*

there exists, on that domain, a function, say $\phi(u, v)$, which is of class C^1 and has the following properties:

(i) *all solutions of the differential equation*

$$dv/du = \phi(u, v) \tag{4}$$

represent geodesic arcs, Γ, of (2) *and*

(ii) *if a geodesic arc, Γ, of* (2) *possesses a tangent at some point and has, at that point, the same direction as a solution path of* (4), *then the latter solution path is identical with the arc Γ.*

Under these assumptions, there exist on (1) *positive, continuous multipliers $\mu = \mu(u, v)$ for the Pfaffian*

$$\omega = (E + F\phi)du + (F + G\phi)dv; \tag{5}$$

in fact, the function

$$\mu = (E + 2F\phi + G\phi^2)^{-\frac{1}{2}} \tag{6}$$

(*which, in view of* (0), *is continuous and positive*) *is such a multiplier.*

In other words, there exists on $\mathcal{B}_a$: $u^2 + v^2 < a^2$ a function $r = r(u, v)$ (unique to an additive constant, say to its value $r(0, 0)$ at the center of $\mathcal{B}_a$) such that r possesses on $\mathcal{B}_a$ the partial derivatives

$$r_u = \mu(E + F\phi), \qquad r_v = \mu(F + G\phi), \tag{7}$$

where μ denotes the function (5).

Since the vanishing of the Pfaffian (5), i. e., the differential equation

$$(E + F\phi)du + (F + G\phi)dv = 0 \tag{8}$$

for $v = v(u)$ or $u = u(v)$, characterizes the transversals of the sheaf of curves defined by the differential equation

$$dv - \phi du = 0, \tag{9}$$

the assertion of Lemma 1 implies that the differential equation of the transversals of the geodesic sheaf defined by (4) possesses *some* (continuous, non-

vanishing) integrating factor, $\mu = \mu(u, v)$. Not even this is obvious, since, if nothing but (0) in Lemma 1 is assumed, and if (5) is written in the form $\omega = M(u, v)du + N(u, v)dv$, where M and N are just continuous, then the Pfaffian ω need not possess a multiplier in any sense; cf. [12], Section 7. Correspondingly (and in view of Section **7** below), it turns out that the content of Lemma 1 is substantially equivalent to the following statement:

LEMMA 2. *Under the assumptions of Lemma* 1 (*and if* a *in* $\mathcal{B}_a$: $u^2 + v^2 < a^2$ *is small enough*), *there exists on* $\mathcal{B}_a$ *a pair of functions,*

$$u^* = u^*(u, v), \qquad v^* = v^*(u, v), \tag{10}$$

such that the transformation (10) *is of class* C^1 *and of non-vanishing Jacobian, and maps the neighborhood* $\mathcal{B}_a$ *of* $(u, v) = (0, 0)$ *on a neighborhood of* $(u^*, v^*) = (0, 0)$ *in such a way that the metric* (2) *acquires a "geodesic" form*

$$ds^2 = du^{*2} + g(u^*, v^*)dv^{*2}, \tag{11}$$

where g *is a positive, continuous function.*

Lemma 1 has a partial converse, as follows:

LEMMA 3. *Suppose that* (2) *satisfies* (0) *in Lemma* 1, *and that a function* $\phi(u, v)$, *which is of class* C^1 *on* $\mathcal{B}_a$, *has the property that the Pfaffian* (5) *possesses the multiplier* (6) *on* $\mathcal{B}_a$. *Then every solution of* (4) *represents a geodesic of* (2).

It remains undecided whether the C^1-assumption imposed on $\phi(u, v)$ in Lemmas 1-3 can be reduced in some extent (for instance, to the extent of requiring only that $\phi(u, v)$ be continuous and such that the solutions of (4), belonging to an initial condition, are unique).

3. *Reduction to parallel segments.* By adapting a scheme from analytical mechanics (in a highly differentiable case), it will be convenient to arrange the proofs in such a way that the sheaf of geodesics which is defined by (4) is assumed to be a sheaf of line segments

$$v = \text{const.}, \tag{12}$$

which means that

$$\phi(u, v) \equiv 0 \tag{13}$$

in (4). It turns out that it can be assumed that, besides (13),

$$F(0, v) \equiv 0 \tag{14}$$

in (2).

In order to make applicable the simplification afforded by this scheme, it will first be proved that, under the assumptions of Lemma 1, 2 or 3, there always exists an admissible (u, v)-transformation of class C^1 which leaves the assumptions unchanged and leads to the normalizations (13)-(14).

Proof. After an affine transformation of (1) (and if a in the new $\mathcal{B}_a$: $u^2 + v^2 < a^2$ is small enough), it can be assumed that the coefficient matrix of (2) is the unit matrix at the center of $\mathcal{B}_a$,

$$(15) \qquad E(0,0) = 1, \qquad F(0,0) = 0, \qquad G(0,0) = 1.$$

Then the function $E + F\phi$ of (u, v), being $1 + 0 \cdot \phi((0,0) > 0$ at the center of $\mathcal{B}_a$, will satisfy

$$(16) \qquad E + F\phi > 0 \text{ on } \mathcal{B}_a,$$

if a is small enough. Hence the differential equation (8) can be written in the form $u' = f(u, v)$, where f is continuous on $\mathcal{B}_a$ and the prime denotes d/dv. Consequently (8) has, for small $|v|$, say for $|v| < b$, at least one solution $u = u(v)$ satisfying $u(0) = 0$. Let $\gamma(v)$ denote such a solution of (8). Thus

$$(17) \qquad u = \gamma(v), \text{ where } -b < v < b \text{ and } \gamma(0) = 0.$$

On the other hand, since $\phi(u, v)$ is of class C^1, the differential equation (9), which is (4), has a unique solution $v = v(u) = v(u; u_0, v_0)$ satisfying $v(u_0) = v_0$, whenever (u_0, v_0) is close enough to $(0, 0)$, and the function $v(u; u_0, v_0)$ exists and is of class C^1 in its three arguments together, if (u_0, v_0) is restricted to a sufficiently small circle about $(0, 0)$, and u to a sufficiently short interval $-c < u < c$ (the length of which is independent of u_0, v_0).

In terms of the function $v(u; u_0, v_0)$ and of the function $\gamma(v)$ occurring in (17), define, for small $|\alpha|$, $|\beta|$, two functions, U and V, by placing

$$(18) \qquad U(\alpha, \beta) = \alpha + \gamma(\beta), \qquad V(\alpha, \beta) = v(\alpha + \gamma(\beta);\ \gamma(\beta), \beta),$$

and consider the transformation

$$(19) \qquad u = U(\alpha, \beta), \qquad v = V(\alpha, \beta).$$

This mapping of a neighborhood of $(\alpha, \beta) = (0, 0)$ on a neighborhood of $(u, v) = (0, 0)$ is of class C^1 (since the functions $\gamma(v)$, $v(u; u_0, v_0)$ occurring in (18) are), and the Jacobian of the transformation (19) does not vanish at, hence near, $(0, 0)$. In fact, the Jacobian $\partial(U, V)/\partial(\alpha, \beta)$ of (18) is $V_\beta - \gamma'\phi$, where

$$V_\beta = \partial V/\partial \beta = \gamma'\phi + \gamma'\partial v/\partial u_0 + \partial v/\partial v_0.$$

Since $\partial v/\partial w_0$ at $(u; u_0, v_0) = (0; 0, 0)$ becomes 0 or 1 according as $w_0 = u_0$ or $w_0 = v_0$, it follows that the Jacobian at $(\alpha, \beta) = (0, 0)$ is $0 + 0 + 1 \neq 0$.

The meaning of the transformation (18)-(19) (which corresponds to the " transformation to the rectilinear motion " in the Hamilton-Jacobi theory) is as follows: For a given (u_0, v_0), the solution path $v = v(u; u_0, v_0)$ of (9) meets the solution path (17) of (8) at a unique point $(\gamma(\beta_0), \beta_0)$. The inverse of the transformation (18)-(19) is $(u_0, v_0) \to (\alpha_0, \beta_0)$, where $\alpha_0 = u_0 - \gamma(\beta_0)$. For a fixed β, the arc (19) is a solution path of (9).

It is seen either from this interpretation or from a direct calculation that, if (e, f, g) is the (α, β)-representation of the covariant tensor (E, F, G), i. e., if (2) is identical with

$$ds^2 = e(\alpha, \beta)\,d\alpha^2 + 2f(\alpha, \beta)\,d\alpha d\beta + g(\alpha, \beta)\,d\beta^2$$

by virtue of the C^1-transformations (18)-(19), then the functions e, f, g are identical with

$$2F\phi + G\phi^2, \qquad E\gamma' + F\gamma' + FV_\beta + G\phi V_\beta, \qquad E\gamma'^2 + 2F\gamma' V_\beta + GV_\beta^2,$$

respectively. It is clear from (18) that (4) is transformed into $d\beta/d\alpha = 0$, which means that (13) will hold if (α, β) is called (u, v). In addition, since the arc $u = \gamma(\beta)$, $v = \beta$ becomes the transversal (17) when $\alpha = 0$, condition (14) will be satisfied if $\alpha, \beta; f$ are called $u, v; F$. Finally, the Pfaffian (5) and the function (6) are transformed into $\omega^* = e\,d\alpha + f\,d\beta$ and $\mu^* = e^{-\frac{1}{2}}$, respectively. Since the property of a Pfaffian to be an exact differential and the property of an arc to be a geodesic are invariant under a local C^1-transformation $(u, v) \to (\alpha, \beta)$ of non-vanishing Jacobian, this proves that the assumptions (13) and (14) will not involve a loss of generality in the proof of Lemmas 1-3.

4. *Proof of Lemma* 1. In view of (13), the differential equations (9) and (8) reduce to

$$dv/du = 0 \tag{20}$$

and

$$du/dv = -F(u, v)/E(u, v), \qquad (E > 0), \tag{21}$$

respectively, and since the function (6) becomes $\mu = E^{-\frac{1}{2}}$, the pair of relations (7) simplifies to

$$(22_1)\quad r_u = E^{\frac{1}{2}}(u, v); \qquad (22_2)\quad r_v = F(u, v)/E^{\frac{1}{2}}(u, v).$$

Hence the assertion of Lemma 1 is that, under the assumptions (0) and (i)-(ii), there exists on $\mathcal{G}_a$ a function $r = r(u, v)$ of class C^1 satisfying both (22_1) and (22_2).

It will be proved that such an $r(u,v)$ is given by

$$(23) \qquad r(u,v)=\int_0^u E^{\frac{1}{2}}(t,v)\,dt.$$

It is clear that the function (23) is continuous and that its partial derivative $r_u(u,v)$ exists and satisfies (22_1) (and is therefore continuous). What is not clear is that the partial derivative $r_v(u,v)$ occurring in assertion (22_2) exists at all; in fact, the function integrated in (23) is just continuous.

With reference to a fixed number r, consider the equation

$$(24) \qquad r(u,v)=r(=\text{const.}).$$

In view of (23), this equation is satisfied at $(u,v)=(0,0)$ if $r=0$. It follows therefore from (22_1), where $E\neq 0$, and from standard facts on implicit functions, that (24) defines, in a neighborhood of $(v;r)=(0;0)$, a unique continuous function

$$(25) \qquad u=u(v;r)$$

in such a way that a triple $(u,v;r)$ sufficiently close to the triple $(0,0;0)$ will satisfy (24) if and only if the u in $(u,v;r)$ is of the form (25). Furthermore, since $E\neq 0$ in (22_1), the function (25), defined by (24), has a continuous partial derivative with respect to r, and this derivative is given by

$$(26) \qquad u_r(u;r)=E^{-\frac{1}{2}}(u,v), \text{ where } u=u(v;r).$$

It will be proved that

(a) *there exists a positive number d having the property that, if r is any fixed value for which $|r|$ is sufficiently small, then the continuous function* (25) *exists for $|v|<d$ and represents a solution of* (21).

If (a) is granted, then the proof of Lemma 1 can be completed as follows: (a) implies that the function (25) has, in a neighborhood of $(v;r)=(0,0)$, a partial derivative with respect to v, and that this derivative satisfies (21), i. e., that

$$(27) \qquad u_v(v;r)=-F(u,v)/E(u,v), \text{ where } u=u(v;r),$$

which implies that this derivative is continuous in v and r together. Since (24) and (25) are equivalent, it follows that the function (23) of (u,v) is of class C^1. Finally, (22_2) follows from (25), (26) and (27).

In order to prove (a), it will first be shown that

(β) *if a point* (u^0, v^0) *is sufficiently near the point* $(u, v) = (0, 0)$, *then the distance,* $\int ds$, *from the point* (u^0, v^0) *to the line* $u = 0$ *is minimized by the geodesic* $v = v^0$.

5. *Proof of* (β). This will be the only part of the proof of Lemma 1 in which the assumption (ii) is used.

First, if (u^0, v^0), where $u^0 \neq 0$, is a point close enough to the center of $\mathcal{E}_a$: $u^2 + v^2 < a^2$, then the existence of a rectifiable Jordan arc, say $\Gamma = \Gamma(u^0, v^0)$, minimizing the distance ($= \int ds$) of (u^0, v^0) from the line $u = 0$, follows by the same procedure as the existence of an arc minimizing the distance between two given points which are close enough (Hilbert). If there are more than one $\Gamma = \Gamma(u^0, v^0)$, choose one of them, and denote by $(0, v^*)$ the point at which this Γ reaches the line $u = 0$. It will be shown that Γ is transversal to the line $u = 0$ at the point $(0, v^*)$. Then the assumption (ii) will assure that Γ is the geodesic $v = v^*$; cf. (12), (13). This in turn will imply that $v^* = v^0$ and that Γ is, therefore, the geodesic $v = v^0$, as claimed by (β).

Suppose if possible that $\Gamma = \Gamma(u^0, v^0)$ is not transversal to the line $u = 0$ at the point $(0, v^*)$. Then Γ either does not have a tangent at $(0, v^*)$ or, if it does, that tangent has a direction which fails to be transversal to $u = 0$ (by a tangent is meant a unilateral tangent, since Γ can be assumed to end at the point $v = v^*$ of the line $u = 0$). Under either of these hypothesis, it is seen from (14) and (21) that there exists on Γ a sequence of points $(u_1, v_1), (u_2, v_2), \cdots$ which converge to the point $(0, v^*)$ and have the property that the inequality

$$(28) \qquad |(v_n - v^*)/u_n| > c,$$

holds for a positive c which is independent of n ($= 1, 2, \cdots$). Since $u^0 \neq 0$, it can be assumed that $u^0 > 0$, and also that (28) is replaced by

$$(29) \qquad v_n - v^* > cu_n > 0, \qquad (n = 1, 2, \cdots)$$

(the other possibilities can be treated similarly).

Accordingly, the proof of (β) will be complete if it is shown that the existence of a $c > 0$ satisfying (29), where $(u_n, v_n) \to (0, v^*)$ as $n \to \infty$, leads to a contradiction. The latter will result from the fact that (29) leads to the following conclusion:

(?) If n is large enough, then the distance $\int ds$ between the two points $(0, v_n)$, (u_n, v_n), when measured along the geodesic $v = v_n$, is shorter than

the distance $\int ds$ between the two points $(0, v^*)$, (u_n, v_n), when measured along $\Gamma = \Gamma(u^0, v^0)$.

In order to deduce the latter assertion, (?), subject the (u, v)-plane to an affine transformation $(u, v) \rightarrow (x, y)$ in such a way that (2) becomes $ds^2 = dx^2 + dy^2$ at the point $(u, v) = (0, v^*)$. Then the distance $\int ds$ along any geodesic which joins any two points is $(1 + o(1))d$ as those two points tend to the point $(0, v^*)$, where d denotes the Euclidean distance in the (x, y)-plane between the two points; cf. [7], pp. 144-148. On the other hand, it is clear that the Euclidean distance between the two points $(0, v^*)$, (u_n, v_n) is not less than $(1 + o(1))/\sin\alpha$ times the Euclidean distance between the two points $(0, v_n)$, (u_n, v_n), if, with reference to the positive number c occurring in (29), the $\alpha = \alpha_c$ in $1/\sin\alpha$, where $0 < \alpha < \frac{1}{2}\pi$, denotes the angle between the (x, y)-images of the two lines $v - v^* = cu$, $u = 0$. Since $1/\sin\alpha > 1$, and since $(u_n, v_n) \rightarrow (0, v^*)$ as $n \rightarrow \infty$, the preceding two o-relations imply the truth of (?).

The proof of (β) is now complete, since (?) contains a contradiction. In fact, if n is large enough, and if Γ_n denotes the path consisting of the portion $0 \leqq u \leqq u_n$ of the geodesic $v = v_n$ and of that portion of $\Gamma = \Gamma(u^0, v^0)$ which joins (u^0, v^0) to (u_n, v_n), then (?) implies that the length $\int ds$ of the path Γ_n, a path joining a point of the line $u = 0$ to the point (u^0, v^0), is shorter than the length $\int ds$ of $\Gamma(u^0, v^0)$. But this contradicts the definition of $\Gamma(u^0, v^0)$.

6. *Proof of* (α). Suppose if possible that the assertion of (α) is false. Then there exists on the arc (24) or (25) a point (u_0, v_0), arbitrarily close to the point $(0, 0)$, in such a way that the arc either does not have a tangent at (u_0, v_0) or, if it does, that tangent has a direction which fails to be transversal to the geodesic $v = v_0$. It can be assumed that $u \neq 0$, say $u_0 > 0$. For, if $r = 0$, then (23) shows that (25) is the arc $u = u(v; 0) \equiv 0$ which, in view of (14), represents a solution of (21).

In order to simplify the notations, subject the (u, v)-plane to an affine transformation after which the metric (2) reduces to $ds^2 = du^2 + dv^2$ at the point (u_0, v_0). Then, in both of the cases negating the truth of (α), there exists on the arc (24) or (25) a sequence of points $(u^1, v^1), (u^2, v^2), \cdots$ which converge to the point (u_0, v_0) and have the property that

$$(30) \qquad | v^n - v_0 | / | u^n - u_0 | < c, \text{ where } u^n \neq u_0 > 0,$$

holds for a positive c which is independent of $n(=1, 2, \cdots)$. Corresponding to the reduction of (28) to (29), it can be assumed that (30) is replaced by

$$(31) \qquad 0 \leqq v^n - v_0 < c(u_0 - u^n) \qquad (n = 1, 2, \cdots).$$

With reference to the constant $c > 0$ occurring in (31), let $\Lambda = \Lambda(u_0, v_0; c)$ denote the line $v - v_0 = (-1/c)(u_0 - u)$. Then it is clear from $(u^n, v^n) \to (u_0, v_0)$ that, if n is large enough, the line Λ will have on the geodesic $v = v^n$ some point, say the point (u_n, v^n).

If the results of [7], pp. 144-148, are applied in the same way as in Section **5** above, it now follows that if n is large enough, the distance $\int ds$ between the two points (u_0, v_0), (u_n, v^n), measured along the line Λ, is shorter than the distance $\int ds$ between the two points (u^n, v^n), (u_n, v^n), measured along the geodesic $v = v_n$. Hence there results a path consisting of a portion of the geodesic $v = v_0$ and of a segment of Λ and having the property that, while it is a path which joins the point $(0, v_0)$ to the point (u_n, v^n), it is shorter than the portion of the geodesic $v = v^n$ which joins $(0, v^n)$ to (u_n, v^n). This contradicts (β) and therefore proves (α).

The proof of Lemma 1 is now complete.

7. *Proof of Lemma* 2. In view of Section **3**, it is sufficient to prove Lemma 2 under the normalizations (13)-(15). Then, according to the proof of Lemma 1, the function (23) is of class C^1 and satisfies (22_1)-(22_2). In terms of this function, define a transformation of a neighborhood of $(u, v) = (0, 0)$ into a neighborhood of $(r, v) = (0, 0)$ by placing

$$(32) \qquad r = r(u, v), \qquad v = v.$$

The Jacobian of the C^1-mapping (32) is $1 \cdot r_u - 0$, which, in view of (22_1), is distinct from 0.

Let the covariant metric tensor of (2) be expressed in terms of the coordinates (32). Then it is readily seen from (22_1)-(22_2) that

$$(33) \qquad ds^2 = dr^2 + (EG - F^2)E^{-1}dv^2$$

is an identity by virtue of (2) and (32). But (33) is of the desired form (11), with $u^* = r$, $v^* = v$, and $g = G - F^2/E$, where the functions E, F, G of u and v are thought of as expressed, by means of the inverse of the substitution (32), as functions of r and v.

8. *Proof of Lemma* 3. In view of Section **3**, it is sufficient to prove Lemma 3 under the assumptions (12)-(15). Then (14) and the assumptions

of Lemma 3 mean that (23) is a function of class C^1 satisfying (22_1)-(22_2).

Section **7** shows that (32) transforms (2) into (33). But (33) is of the form

$$(34) \qquad ds^2 = dr^2 + g(r, v)\,dv^2, \qquad \text{where } g(0,0) = 1,$$

by (15), and it is clear from (34) that, if $r^2 + v^2$ is small enough, every segment (12) minimizes the distance $\int ds$ between any two points of that segment. This proves Lemma 3.

9. *Normal coordinates.* Suppose that a metric (2), satisfying (3), is given in terms of normal coordinates (u, v) at $(0, 0)$ (Riemann). By this is meant that

$$(35) \qquad u = r\cos\theta, \qquad v = r\sin\theta$$

represents a geodesic for every fixed θ and that, in addition, the $r \geqq 0$ occurring in (35) is, except for a factor which depends only on θ, identical with the arc length on each of the geodesics $\theta = \text{const.}$ which issue from $(u, v) = (0, 0)$. *If the functions $E(u, v)$, $F(u, v)$, $G(u, v)$ possess partial derivatives of a sufficiently high order*, then, according to Gauss [5], pp. 249-250 (and, in the multi-dimensional case, Riemann [11], p. 279; cf. Dedekind's comments [4], pp. 406-407), the differential equations of the geodesics (that is, the equations of motion

$$(36) \qquad [L]_u = 0, \qquad [L]_v = 0,$$

where the brackets denote the Lagrangian derivatives of

$$(37) \qquad L(u, v; u', v') = \tfrac{1}{2}E(u, v)u'^2 + F(u, v)u'v' + \tfrac{1}{2}G(u, v)v'^2 = \tfrac{1}{2}s'^2$$

and the prime denotes d/dt) must possess the invariant relations

$$(38) \qquad L_{u'}(u, v; u, v) = L_{u'}(0, 0; u, v), \qquad L_{v'}(u, v; u, v) = L_{v'}(0, 0; u, v)$$

whenever (u, v) is a normal coordinate system at $(0, 0)$, and, as pointed out by Herglotz ([8], p. 216), this necessary condition for normal coordinates is sufficient as well.

Clearly, neither the definition of coordinates which are normal coordinates nor the pair of relations (38) (which, in view of (37), simply mean that

$$(39) \qquad E(u, v)u + F(u, v)v = u, \qquad F(u, v)u + G(u, v)v = v,$$

if, without loss of generality, the normalization (15) is used) involves anything like the italicized hypothesis, that concerning a sufficiently high degree of smoothness of the coefficient functions of (2). Correspondingly, as a relevant

illustration of the above theory of geodesic fields in a metric which is just continuous, it will now be shown that the criterion holds without *any* differentiability assumption on the coefficient functions of (2).

(*) *Suppose that* $E(u,v)$, $F(u,v)$, $G(u,v)$ *are continuous functions satisfying* (3) *in a neighborhood of* $(0,0)$, *and that they are normalized at* $(0,0)$ *by* (15). *Then the pair of identities* (39) *is necessary and sufficient in order that the coordinates* u, v *be normal at* $(u,v)=(0,0)$.

The proof of the necessity and sufficiency of (39) could be deduced from Lemma 3 and Lemma 1, respectively. Corresponding to the circumstance that Lemmas 1-3 and (*) deal with "geodesic parallel" and "geodesic polar" coordinates, respectively, such a deduction of (*) would, however, lead to a complication, since what represents (4) in (*) is the differential equation $dv/du = v/u$, which has a singularity at $u=0$ (corresponding to the vanishing of the Jacobian of (35) at $r=0$). Because of this formal complication, it will be just as convenient to prove (*) directly.

10. *Proof of* (*). A direct substitution shows that (35) transforms (2) into

$$ds^2 = e(r,\theta)\,dr^2 + 2f(r,\theta)\,dr d\theta + g(r,\theta)\,d\theta^2, \tag{40}$$

where, in (binary) vector and matrix notations,

$$\begin{pmatrix} e \\ f/r \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} E & F \\ F & G \end{pmatrix}\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \tag{41}$$

while

$$g/r^2 = E\sin^2\theta - 2F\sin\theta\cos\theta + G\cos^2\theta. \tag{42}$$

In order to prove the sufficiency of (39), suppose that the coordinates u, v satisfy (39). Then (35) shows that (41) can be written in the form

$$\begin{pmatrix} e \\ f/r \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

(where $1 = \cos^2\theta + \sin^2\theta$). Since this means that $e \equiv 1$ and $f \equiv 0$, it follows that (40) reduces to

$$ds^2 = dr^2 + g(r,\theta)\,d\theta^2 \qquad (g>0). \tag{43}$$

But (43) makes it trivial that the distance $\int ds$ between the point $r=0$ and any point (r_0,θ_0), where $r_0>0$, is minimized by the path $\theta = \text{const.}\ (=\theta_0)$, and that r is the arc length along the path. This means that the coordinates (41) are normal at $(0,0)$, as claimed by the second assertion of (*).

In order to prove the necessity of (39), which is the first assertion of (*), it is sufficient to show that $e = 1$ and $f = 0$ are identities in (u, v) if the coordinates u, v are normal at $(0, 0)$. In fact, if the constants $e = 1$, $f = 0$ are substituted into (41), then (41) reduces to a pair of identities which, in view of the definition (35), is precisely (39).

According to the first line of (41) and the normalization (15), the value of $e(u, v)$ at $(u, v) = (0, 0)$ is $\cos^2\theta + \sin^2\theta = 1$. On the other hand, since the coordinates u, v are supposed to be normal at $(0, 0)$, the equations (35) represent, for every fixed θ, a geodesic on which the arc length, when measured from $(0, 0)$, is of the form $s = cr$, where the positive number $c = c(\theta)$ is independent of r. Hence it is clear from (40) that $e(u, v)$ is independent of (u, v). In view of $e(0, 0) = 1$, this proves that $e(u, v)$ is the constant 1.

Consequently, only the identical vanishing of f remains to be proved. But (40) shows that the identical vanishing of f is equivalent to the statement [2] that every arc $r =$ Const. is transversal to every geodesic $\theta =$ const. Suppose if possible that this transversality fails to take place somewhere, i. e., that there exist an $r_0 > 0$ and a $\theta = \theta_0$ for which the arc $r = r_0$ is not transversal to the geodesic $\theta = \theta_0$ at the point (r_0, θ_0). Then, by using results of [7], pp. 144-148, in the same way as above (Section **5**), it follows that there exists a sequence of points (r_n, θ_n) which tend, as $n \to \infty$, to (r_0, θ_0) and possess the following property: Whenever n is large enough, the point (r_n, θ_n) can be joined to the origin $(r = 0)$ by an arc along which the length $\int ds$ is less than r_n. This contradicts, however, the assumption, according to which the arc $\theta = \theta_n$ joining the point (r_n, θ_n) to the point $r = 0$ is a geodesic arc of length r_n.

11. *Jacobi's multiplier.* While Lemmas 1-3 concern themselves with the equation (8) of the transversals of the geodesics defined by (9), the following lemma deals with (9) itself (but assumes that the function ϕ of u and v contains a parameter also).

LEMMA 4. *Let $\phi = \phi(u, v; w)$ be a function of class C^1 on the product space of a sufficiently small domain $\mathcal{B}_a$: $u^2 + v^2 < a^2$ and of an interval $|w| < b$, and suppose that the functions E, F, G and ϕ satisfy assumptions* (0) *and* (i)-(ii) *of Lemma* 1, *when w is fixed. Then, for fixed w, the continuous function*

$$\lambda = (EG - F^2)(E + 2F\phi + G\phi^2)^{-3/2}\phi_w \tag{44}$$

---

[2] This is the analogue of assertion (α), Section **4**, in the proof of Lemma 1. Nothing like assertion (β), Section **4**, is involved in the present case, since the arc $u = 0$, occurring in (β), now degenerates to the point $r = 0$.

of (u, v) represents a multiplier of (9). What is somewhat more, there exists a function $R = R(u, v; w)$ which is continuous on the product space of $\mathcal{B}_a$ and $|w| < b$, is of class C^1 on $\mathcal{B}_a$, and satisfies the relations

$$R_u = -\lambda\phi, \qquad R_v = \lambda. \tag{45}$$

(The vanishing of the partial derivative ϕ_w, and therefore that of the multiplier (44), is not excluded.)

If $E, F, G,$ instead of being just continuous, are sufficiently smooth, then Lemma 4 reduces to Jacobi's theorem concerning a "last multiplier" (in case of geodesics); cf. [10], p. 498 and [3], vol. II, p. 431.

Lemma 4 will be essential in proving, without the usual assumptions of differentiability, a fundamental theorem on non-euclidean geometries; cf. Section **12** below.

Proof of Lemma 4. If w is fixed, then, according to assertions (5)-(6) of Lemma 1, the function

$$r(u, v; w) = \int_{(0,0)}^{(u,v)} \mu\{(E + F\phi)du + (F + G\phi)dv\} \tag{46}$$

satisfies (7) on $\mathcal{B}_a$. Since $\phi(u, v; w)$ is supposed to be of class C^1 (with the inclusion of w), the function (6) has the continuous partial derivative $\mu_w = -(F + G\phi)\mu^3\phi_w$. Hence, if (46) is differentiated with respect to w, a straightforward calculation gives

$$r_w(u, v; w) = \int_{(0,0)}^{(u,v)} (\lambda du - \lambda\phi dv), \tag{47}$$

if λ is defined by (44). Since (47) means that (45) is satisfied by the function

$$R(u, v; w) = r_w(u, v; w), \tag{48}$$

the assertions of Lemma 4 follow.

12. *Beltrami's theorem.* If all geodesic arcs of a metric (2) are segments of straight lines in a domain of some (u, v)-plane, then the metric is of constant curvature. This is a classical theorem of Beltrami ([1], pp. 262-280; cf. [3], vol. III, pp. 41-47). His proof and its variants assume, however, that the coefficient functions of (2) have a sufficiently high degree of differentiability in (u, v). Without such an assumption, the curvature K of

(2) cannot even be defined, and the proofs need substantially stronger restrictions of differentiability than what suffices for equating K to the differential operator assigned by the Theorema Egregium. On the other hand, in view of the fundamental *geometrical* significance of Beltrami's result, it seems to be essential that the theorem should be formulated and proved without any assumption of differentiability, as follows:

(**) *Suppose that the coefficients of* (2) *are continuous functions satisfying* (3) *on a* (u, v)*-domain, say on* $\mathfrak{E}_a$: $u^2 + v^2 < a^2$, *and that every sufficiently short segment* (*of a straight line*, $c_1 u + c_2 v = c_0$) *contained in* $\mathfrak{E}_a$ *is a geodesic arc of* (2). *Then the* (*given*) *coefficients of* (2) *are analytic functions of* (u, v) *on* $\mathfrak{E}_a$, *and the curvature* $K = K(u, v)$ *of* (2) *is independent of* (u, v) *on* $\mathfrak{E}_a$.

As mentioned in Section **1**, a point and a direction do not in general determine a (unique) geodesic of a metric (2), not even if the functions E, F, G of (u, v) are of class C^1 (which is not assumed in the present case). Correspondingly, a substantial part of the proof of the general formulation (**) of Beltrami's theorem will consist in showing that this and similar possibilities are excluded by the last assumption of (**). To this end, the following lemma will be needed:

(†) *Under the assumption of* (**), *every sufficiently short geodesic arc of* (2) *is a segment* (*of a straight line*, $c_1 u + c_2 v = c_0$) *in* $\mathfrak{E}_a$.

The proof of the latter assertion, (†), will depend on the steps used in the proof of Lemma 1.

13. *Proof of* (†). Let $\Gamma: (u = u(t), v = v(t))$, where $0 \leqq t \leqq 1$, be a geodesic of (2) joining a point, say (u_0, v_0), of $\mathfrak{E}_a$ to another point of $\mathfrak{E}_a$. Without loss of generality, the latter point can be assumed to be the origin, $(0, 0)$, since a is meant to be sufficiently small. The assertion of (†) is that the arc Γ must be on a straight line, (35), where $\theta = \theta(u_0, v_0)$ is constant on Γ, and $r = r(t) \geqq 0, 0 \leqq t \leqq 1$.

With reference to an interior point t^* of the given parameter range $0 \leqq t \leqq 1$ of Γ, let Γ_1, Γ_2 denote the respective portions $0 \leqq t \leqq t^*$, $t^* \leqq t \leqq 1$ of the given geodesic, $\Gamma = \Gamma_1 + \Gamma_2$, and let Γ^1, Γ^2 be the segments (of straight lines) which join the point $P^* = (u(t^*), v(t^*))$ of Γ to its respective points $(u(0), v(0)) = (0, 0)$, $(u(1), v(1))$. Then Γ^i, where $i = 1, 2$, has the same length $\int ds$ as Γ_i. For, on the one hand, the assumptions of (†), being those of (**), imply that the segment Γ^i minimizes the distance $\int ds$ between its given end points and, on the other hand, the same is true of the geodesic

arc Γ_i, the end points of which are the same as those of Γ^i. Consequently, $\Gamma_1 + \Gamma^2$ and $\Gamma^1 + \Gamma_2$ are of the same length, and join the same points as the geodesic $\Gamma = \Gamma_1 + \Gamma_2$, as well as the polygonal path $\Gamma^1 + \Gamma^2$, and are therefor geodesics.

Since $\Gamma^1 + \Gamma^2$ is a geodesic containing the point $P^* = (u(t^*), v(t^*))$, an application of arguments applied in [7], pp. 144-148 (or, equivalently, in Section **5** above) shows that $\Gamma_1 + \Gamma^2$ must have at P^* a unilateral tangent line from the "right," and that this line is that containing the segment Γ^1. Since the latter determines for $\Gamma_1 + \Gamma^2$ a unilateral tangent line from the "left," it now follows from Corollary 1 in [7], p. 145, that $\Gamma_1 + \Gamma^2$ must have at P^* a tangent, i. e., that the two unilateral tangents coincide at P^*. For reasons of symmetry, the same is true of the geodesic $\Gamma^1 + \Gamma_2$, as well as of the geodesic $\Gamma^1 + \Gamma^2$.

Clearly, this is possible only if the two segments Γ^1, Γ^2 issuing from P^* are collinear, and if the straight line containing them represents a (bilateral) tangent of $\Gamma_1 + \Gamma_2 = \Gamma$ at P^*. Since the sum of the segments Γ^1, Γ^2 is a segment, $\Gamma^1 + \Gamma^2$ which is a geodesic joining the points $(u(0), v(0))$, $(u(1), v(1))$, points which are independent of the choice of $P^* = (u(t^*), v(t^*))$ on $\Gamma = \Gamma_1 + \Gamma_2$, it follows that Γ must be identical with the segment $\Gamma^1 + \Gamma^2$. This proves assertion (†) of Section **12**.

14. *Field constructions for* (**). Let w be any constant, and let

$$\phi = \phi(u, v; w) \equiv w \tag{49}$$

for every (u, v) on (1). The assumptions of (**) show that assumptions (0) and (i) of Lemma 1 are satisfied by the case (49) of (4) (for every fixed w). On the other hand, (†) shows that assumption (ii) of Lemma 1 is satisfied. Consequently, Lemma 1 is applicable (at every fixed value of w), as is Lemma 4 (the function (49) is of class C^1 in $u, v; w$ together).

Since the case $w = 0$ of (49) reduces (5) and (6) to $Edu + Fdv$ and $E^{-\frac{1}{2}}$ respectively, it follows from Lemma 1 that

$$E^{-\frac{1}{2}}(Edu + Fdv) \text{ is an exact differential.} \tag{50}$$

On the other hand, it is seen from the case (49) of the definition (44) of λ, and of the assertion (45) of Lemma 4, that since $\phi_w = w_w = 1$, the Pfaffian

$$(EG - F^2)(E + 2Fw + Gw^2)^{-3/2}(dv - wdu)$$

is an exact differential for every value of the constant w. Hence the choice $w = 0$ leads to the conclusion that

$$(EG - F^2)E^{-3/2} \text{ is a function of } v \text{ only.} \tag{51}$$

In addition, since the assumptions of (**), from which (50) and (51) have been concluded, remain unaltered if u, v and E, G are replaced by v, u and G, E respectively, it is clear that, corresponding to (50) and (51),

$$(52)\qquad G^{-\frac{1}{2}}(F\,du + G\,du) \text{ is an exact differential,}$$

and

$$(53)\qquad (EG - F^2)G^{-3/2} \text{ is a function of } u \text{ only.}$$

Beltrami's proof of the assertion, $K = \text{const.}$, of his theorem falls into two parts. He first concludes ([1], pp. 263-266) the preceding four relations, (50)-(53), under the assumption that the functions E, F, G are sufficiently smooth (of class C^n, with something like $n = 2$), and then, by assuming a still higher degree of differentiability (something like $n = 5$), he deduces ([1], pp. 266-270) the assertion, $K = \text{const.}$, from (50)-(53). Correspondingly, the completion of the proof (**) will depend on a suitable adaptation of the latter part of Beltrami's proof, leading from (50)-(53) to certain functional equations for which it turns out that their non-analytic solutions cannot be continuous (or, for that matter, L-integrable).

15. *The functional equations of* (**). Let $U = U(u)$, $V = V(v)$ denote, for sufficiently small $|u|$, $|v|$, the cube roots of the respective functions (53), (51). In view of (3), these continuous functions of the respective single variables u, v are positive. Since (53) and (51) mean that both products $UG^{\frac{1}{2}}$, $VE^{\frac{1}{2}}$ are identical with the cube root of $EG - F^2$, it follows that there exists a positive, continuous function $\lambda = \lambda(u, v)$ satisfying

$$(54)\qquad E^{\frac{1}{2}} = \lambda U, \qquad G^{\frac{1}{2}} = \lambda V,$$

and that a continuous function $\mu = \mu(u, v)$ is, therefore, defined by placing

$$(55)\qquad F = \mu\lambda UV$$

(the multipliers (44), (6) have nothing to do with the present λ, μ). In terms of (54) and (55), the remaining two relations, (50) and (52), mean that both Pfaffians

$$(56)\qquad \lambda U du + \mu V dv, \qquad \mu U du + \lambda V dv \quad \text{are exact differentials.}$$

Since U, V are positive, continuous functions of the respective single variables u, v, the conditions

$$(57)\qquad d\alpha = U du + V dv, \qquad d\beta = U du + V dv$$

and $\alpha(0, 0) = 0, \beta(0, 0) = 0$ define, near $(u, v) = (0, 0)$, a pair of functions

$\alpha = \alpha(u, v)$, $\beta = \beta(u, v)$ which are of class C^1 and of non-vanishing Jacobian (the latter being $\partial(\alpha, \beta)/\partial(u, v) = -2UV < 0$). Thus $(u, v) \to (\alpha, \beta)$ and $(\alpha, \beta) \to (u, v)$ are one-to-one transformations, of class C^1, of corresponding small neighborhoods of the origins of the parameter planes (u, v), (α, β). Since $U(u)$ and $V(v)$ are positive and continuous, it follows that, in a neighborhood of $(\alpha, \beta) = (0, 0)$, two positive continuous functions, A and B, of single variables are defined by placing

$$(58) \qquad 1/U(u) = A(\alpha + \beta), \qquad 1/V(v) = B(\alpha - \beta).$$

In fact, since (57) means that $d(\alpha + \beta) = Udu$ and $d(\alpha - \beta) = Vdv$, the functions u, v of (α, β) depend only on $\alpha + \beta$, $\alpha - \beta$ respectively.

According to (56), both Pfaffians

$$(\lambda + \mu)(Udu + Vdv), \qquad (\lambda - \mu)(Udu - Vdv)$$

are exact differentials. In view of (57), this means that the same is true of both Pfaffians $(\lambda + \mu)d\alpha$, $(\lambda - \mu)d\beta$ (in which the functions λ, μ of (u, v) are thought of as expressed in terms of the new variables, α and β). Consequently, the continuous functions $\lambda + \mu$, $\lambda - \mu$ of (α, β) depend only on α, β respectively. Thus if $2a$, $2b$ denote these continuous functions of the respective single variables α, β, then

$$(59) \qquad \lambda = a(\alpha) + b(\beta), \qquad \mu = a(\alpha) - b(\beta).$$

Hence it is seen from the definitions of U and V, (54), (55) and (58), that

$$(60) \qquad a(\alpha) + b(\beta) = 4a(\alpha)b(\beta)A(\alpha + \beta)B(\alpha - \beta).$$

As pointed out above, λ and A, B are positive. It follows therefore from the first of the relations (59) and from (60) that $a > 0$, $b > 0$. In particular, division of (60) by $ab \neq 0$ is allowed; so that

$$(61) \qquad 1/a(\alpha) + 1/b(\beta) = 4A(\alpha + \beta)B(\alpha - \beta).$$

If $\alpha + \beta$ and $\alpha - \beta$ in (61) are replaced by 2α and 2β, respectively, then an integration with respect to the new β leads to the following identity in (α, β):

$$\int_{\alpha}^{\alpha+\beta} dt/a(t) - \int_{\alpha}^{\alpha-\beta} dt/b(t) = 4A(2\alpha)\int_{0}^{\beta} B(2t)\,dt.$$

Since the sum on the left of this identity has a continuous partial derivative with respect to α, the same is true of the product on the right. This means that $A(\alpha)$ has a continuous first derivative. If the rôles of the original α and

β are interchanged in this deduction, it follows that $B(\beta)$ has a continuous first derivative. Finally, a repetition of this argument shows that $A(\alpha)$ and $B(\beta)$ have derivatives of arbitrarily high order.

Consequently, the tacit assumptions of the caluculations of Beltrami ([1], pp. 266-272), referred to at the end of Section **14** above, are satisfied by necessity. In other words, the classical proofs of Beltrami's theorem can now be repeated; so that the proof of (**), Section **12**, is complete.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] E. Beltrami, *Opere Matematiche*, vol. 1 (1902).

[2] O. Bolza, *Vorlesungen über Variationsrechnung*, Leipzig und Berlin, 1909.

[3] G. Darboux, *Leçons sur la théorie générale des surfaces*, Paris, vol. II (1889) and vol. III (1894).

[4] R. Dedekind, see [11].

[5] C. F. Gauss, *Werke*, vol. 4 (1873).

[6] J. Hadamard, *Leçons sur le calcul des variations*, Paris, 1910.

[7] P. Hartman and A. Wintner, "On the problems of geodesics in the small," *American Journal of Mathematics*, vol. 73 (1951), pp. 132-148.

[8] G. Herglotz, Zur Riemannschen Metrik, Berichte der Sächsischen Akademie der Wissenschaften zu Leipzig, vol. 73 (1921), pp. 215-225.

[9] D. Hilbert, "Ueber das Dirichlet'sche Prinzip," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 8 (1900), pp. 184-188.

[10] C. G. Jacobi, *Gesammelte Werke*, vol. 4 (1886).

[11] B. Riemann, *Gesammelte mathematische Werke* (ed. 1892).

[12] A. Wintner, "On isometric surfaces," *American Journal of Mathematics*, vol. 74 (1952), pp. 194-214.

NOTE ON DOUBLE-MODULES OVER ARBITRARY RINGS.*

By Tadasi Nakayama.

Jacobson's [3] module-theoretical Galois theory of non-normal extension fields was generalized by Hochschild [2] to a theory of double-modules over sfields. The theory was further extended to the infinite-dimensional case by Dieudonné [1]. On the other hand, it was shown in [4] that a goodly portion of the theory can be transferred to the case of general rings. It is now natural to study the infinite-dimensional case also for general rings, which we propose to do in this note. It seems to the writer that there are [1] two main features in the theory; one is the characterization of relation-modules, and the other is the characterization of direct (Kronecker) self-products. With respect to the latter, our generalization is rather satisfactory (§ 5), while it is quite powerless with respect to the former; the same was the case with [4]. In closing the introduction, we remark that the formulation of the present note is left-right symmetric to the one in [4] (but is in accord with [1], [2]).

1. Relation-modules of double-modules. Let K be a ring; by a ring we mean in this note always one which posesses a unit element, by a subring we mean one containing that unit element, and by a module, either left- or right-, we mean one on which the unit element operates as an identity. Let A be a second ring, and $\mathfrak{A}$ the additive group of all additive homomorphisms of A into K. $\mathfrak{A}$ is an A-K-double-module with respect to the natural operations defined by [2]

$$x(a\cdot\alpha)=(xa)\alpha,\qquad x(\alpha\cdot k)=(x\alpha)k \tag{1}$$

$$(x, a\,\varepsilon\, A;\, k\,\varepsilon\, K;\, \alpha\,\varepsilon\,\mathfrak{A}).$$

Let M be a K-A-double-module, and let M^* be the module of all K-homomorphisms of M into K, i. e. the dual module of the K-(left-)module M. It is an A-K-double-module according to a definition similar to the above.

* Received May 31, 1951.

[1] Besides the duality between certain subrings (over which the whole ring has independent (right-) bases) and certain relation-modules.

[2] The operations of the elements of A, K on $\mathfrak{A}$ are indicated by dots.

(In fact, the above construction depends only on the fact that A is an A-right-module and K is a K-right-module.) Let u_0 be an element of M. With $\sigma \varepsilon M^*$ we consider the element $\bar{\sigma}$ of $\mathfrak{A}$ defined by

$$x\bar{\sigma} = (u_0x)\sigma \qquad (x \varepsilon A). \tag{2}$$

Denote the totality of $\bar{\sigma}$'s (σ running over M^*), by $R(M, u_0)$. It is an A-K-submodule of $\mathfrak{A}$ as is readily seen, and we call it the *relation-module* of u_0 in M. We have evidently

LEMMA 1. *If M is contained in a A-K-double-module M_1, as K-A-module, such that every element of M^* can be extended to an element of $M_1{}^*$, then $R(M_1, u_0) = R(M, u_0)$.*

Let N be a second K-A-module, and v_0 an element of N. Let ϕ be a (K-A-)homomorphism of M into N. It induces in a natural manner an (A-K-)homomorphism $\phi^* : \tau \rightarrow \sigma = \phi\tau$ of N^* into M^*. Suppose now $u_0\phi = v_0$. Let $\bar{\tau}$ in $R(N, v_0)$ be given by $x\bar{\tau} = (v_0x)\tau$. Then

$$x\bar{\tau} = (v_0x)\tau = ((u_0\phi)x)\tau = ((u_0x)\phi)\tau = (u_0x)\sigma = x\bar{\sigma}. \tag{3}$$

Thus we have the following proposition, whose latter half includes Lemma 1.

PROPOSITION 1. *If there exists a (K-A-)homomorphic mapping ϕ of M into N which maps u_0 into v_0, then*

$$R(M, u_0) \supseteq R(N, v_0). \tag{4}$$

If ϕ^* maps N^* onto M^* then $R(M, u_0) = R(N, v_0)$.

We can immediately verify

PROPOSITION 2. *If $u_0 \varepsilon M$, $v_0 \varepsilon N$, then*

$$R(M \oplus N, w_0) = R(M, u_0) + R(N, v_0) \tag{5}$$

with $w_0 = u_0 + v_0$ in $M \oplus N$.

Let S be a third ring, and let N be an S-K-double-module (contrary to the above). The direct product $N \times M = N \times_K M$ of N, M over K is defined as usual, and is an S-A-double-module. With $\sigma \varepsilon M^*$, $\tau \varepsilon N^*$ (where N^* is the dual module of the S-module N), we set

$$(v \times u)(\sigma \times \tau) = (v(u\sigma))\tau, \tag{6}$$

and observe that this essentially defines $\sigma \times \tau$ as an element of $(N \times M)^*$, independent of the special expression of $v \times u$. Let $\bar{\sigma}$, $\bar{\tau}$ be the elements of

$R(M, u_0)$, $R(N, v_0)$ (v_0 being an element of N), which correspond to σ, τ, respectively. We have

$$(x\bar{\sigma})\bar{\tau} = ((u_0x)\sigma)\bar{\tau} = (v_0((u_0x)\sigma))\tau$$
$$= (v_0 \times (u_0x))(\sigma \times \tau) = ((v_0 \times u_0)x)(\sigma \times \tau).$$

Hence $x \rightarrow (x\bar{\sigma})\bar{\tau}$ is the element of $R(N \times M, v_0 \times u_0)$ corresponding to $\sigma \times \tau \varepsilon (N \times M)^*$, and we have

Proposition 3. *For the product* $N \times M$,

$$R(M, u_0)R(N, v_0) \subseteq R(N \times M, v_0 \times u_0). \tag{7}$$

Furthermore we have

Proposition 4. *Let M be a K-A-double-module generated, as K-A-module, by a single element u_0. Let N be a second K-A-double-module such that for every non-zero element v of N, there exists at least one element τ of N^* satisfying $v\tau \neq 0$. Then the inclusion (4) implies, conversely, the existence of a (unique) homomorphism ϕ of M into N, such that $u_0\phi = v_0$.*

For a proof we observe, after Hochschild [2], that if a certain sum $\Sigma\, ku_0a$ ($k \varepsilon K, a \varepsilon A$) vanishes, then $\Sigma\, k(a\bar{\sigma}) = \Sigma\, k((u_0a)\sigma) = (\Sigma\, ku_0a)\sigma = 0$. If (4) is the case, then this implies $\Sigma\, k(a\bar{\tau}) = 0$, or $(\Sigma\, kv_0a)\tau = 0$, for all $\tau \varepsilon N^*$. This implies in turn $\Sigma\, kv_0a = 0$, according to our assumption on N^*. Thus $u_0 \rightarrow v_0$ defines a (K-A-)homomorphism of M into N.

2. Restricted relation-modules. So far, no particular assumptions have been made on M (nor on K, A). Let us now assume that M possesses an independent K-basis, say $\{u_h\}$. Let σ_h be the element of M^* such that

$$u_i\sigma_h = \delta_{ih} \qquad \text{(Kronecker } \delta\text{'s)}; \tag{8}$$

$u\sigma_h$ is nothing but the coefficient of u_h in the K-linear expression of u by $\{u_h\}$. The K-combinations of σ_h's (with varying h) form a K-submodule $M^\#$ of M^*, which is independent of the particular choice of the basis $\{u_h\}$. Letting σ run only over $M^\#$, we then obtain a submodule $R^\#(M, u_0)$ of $R(M, u_0)$, which we shall call the *restricted relation-module* of u_0 in M. If $\bar{\sigma}_h$ is obtained from σ_h, then $\{\bar{\sigma}_h\}$ forms a (not necessarily independent) K-basis of $R^\#(M, u_0)$. Unlike $R(M, u_0)$, the restricted relation-module $R^\#(M, u_0)$ is not A-left-allowable in general.

Similarly to Lemma 1 we have

Lemma 2. *If M is contained in a K-A-double-module M_1 as K-A-module,*

and M_1 has an independent K-basis which contains an independent K-basis of M, then $R^\#(M_1, u_0) = R^\#(M, u_0)$.

Next let M be homomorphically mapped into N by ϕ, and $u_0\phi = v_0$. Suppose also N has an independent K-basis, say $\{v_j\}$. Expressing each $u_h\phi$ by $\{v_j\}$, we see that for $\tau \varepsilon N^\#$, the element $\phi\tau$ of M^* is in $M^\#$. It follows readily that $R^\#(N, v_0) \subseteq R^\#(M, u_0)$. That Proposition 2 can be transferred to restricted relation-modules is trivial. Thus

PROPOSITION 5. *Our Propositions* 1, 2 *hold also for the restricted relation-modules $R^\#$ (instead of R), provided that both M, N possess independent K-bases.*

As for Proposition 3, we observe that if $\{u_h\}$, $\{v_j\}$ are, respectively, an independent K-basis of M and an independent S-basis of N, then $\{v_j \times u_h\}$ is an independent S-basis of $N \times M = N \times_K M$. Also

$$(v_j \times u_h)(\sigma_i \times \tau_k) = (v_j(u_h\sigma_i))\tau_k = \delta_{hi}\delta_{jk}. \tag{9}$$

On the other hand, the element of $R^\#(N \times M, v_0 \times u_0)$ corresponding to $\sigma_i \times \tau_k$, is just the product $\bar{\sigma}_i\bar{\tau}_k$, as was seen formerly. Thus we have

PROPOSITION 6. *Let M, N be K-A-, and S-K-double-modules, respectively, possessing independent K-, S-bases. Then*

$$R^\#(M, u_0)R^\#(N, v_0) \supseteq R^\#(N \times M, v_0 \times u_0). \tag{10}$$

We have furthermore, corresponding to Proposition 4,

PROPOSITION 7. *Let M, N be K-A-double-modules possessing independent K-bases. If M is generated by u_0 as a K-A-module, and if $R^\#(M, u_0) \supseteq R^\#(N, u_0)$, then $u_0 \rightarrow v_0$ gives a (K-A-)homomorphism of M into N.*

Let us next consider a K-A-double-module M which (not only is generated by u_0 and possesses an independent K-basis, but) possesses an independent K-basis consisting of elements contained in u_0A; then we say that M is a *special K-A-double-module*, and u_0 is a generator. We prove

PROPOSITION 8. *If M is a special K-A-double-module with generator u_0, and $\{u_h\}$ is anarbitrary independent K-basis of M (not necessarily contained in u_0A), then $\{\bar{\sigma}_h\}$ (with σ_h as in* (8)) *forms an independent K-(right-)basis of $R^\#(M, u_0)$.*

It suffices to consider the case where $\{u_h\}$ is contained in u_0A. Let

$$u_h = u_0 t_h \qquad (t_h \varepsilon A). \tag{11}$$

Then

$$t_h \bar{\sigma}_i = \delta_{hi}. \tag{12}$$

It follows immediately that the $\bar{\sigma}_i$'s are independent. That they form a K-basis of $R^{\#}(M, u_0)$ has been seen before.

3. Direct self-products of rings. We now consider the case $K = A$. Then $\mathfrak{A}$ is the absolute module-endomorphism ring of A. Let S be a subring of A. The direct self-product $M = A \times {}_S A$ of A over S is an A-double-module (i. e. A-A-double-module), and we have

PROPOSITION 9. *The relation-module $R(M, 1 \times 1)$ of $M = A \times {}_S A$ with respect to 1×1, is the S-left-endomorphism ring of A (i. e. the commuter in $\mathfrak{A}$ of the left-multiplication ring S_L of S on A), denoted by $E(A, S_L)$.*

For, with $\bar{\sigma} \varepsilon R(M, 1 \times 1)$, where $\sigma \varepsilon M^*$, we have $x\bar{\sigma} = ((1 \times 1)x)\sigma = (1 \times x)\sigma$. For $s \varepsilon S$ we have

$$(sx)\bar{\sigma} = (1 \times sx)\sigma = (s \times x)\sigma = (s(1 \times x))\sigma = s((1 \times x)\sigma) = s(x\bar{\sigma}),$$

which shows that $\bar{\sigma} \varepsilon E(A, S_L)$. Conversely, if $\alpha \varepsilon E(A, S_L)$, we put $(x \times y)\sigma = x(y\alpha)$, and observe that this defines σ uniquely as an element of M^*, since $(xs \times y)\sigma = (xs)(y\alpha) = x(sy\alpha) = (x \times sy)\sigma$. Clearly $\alpha = \bar{\sigma}$.

Now let M be an arbitrary A-double-module and u_0 an element of M. Then the set of all x satisfying $au_0 = u_0 a$ forms a subring of A, and we take it as S:

$$S = \{x \varepsilon A \mid xu_0 = u_0 x\}. \tag{13}$$

Then we have $R(M, u_0) \subseteq E(A, S_L)$; the proof is similar to the first half of the above proofs of Proposition 9 (which depended only on $(1 \times 1)s = s(1 \times 1)$). In other words, if $s \varepsilon S$, then the left-multiplication s_L (on A) commutes with every element of $R(M, u_0)$. The converse is also true if M is such that for every non-zero element u in M, there exist a σ in M^* satisfying $u\sigma \neq 0$.

Now we wish to know when M is isomorphic to $A \times {}_S A$. To do so, let us assume that M is special, and u_0 is its generator. Let $\{u_h\}$ be an independent A-left-basis of M contained in $u_0 A$, and put $u_h = u_0 t_h$ (i. e. (11)). We have then (12). Let N be a second (special) A-double-module which is isomorphic to M. Let v_0, v_h be its elements corresponding to u_0, u_h. Suppose that $u_0 \to v_0 \times u_0$ gives an (A-two-sided) homomorphism of M into $N \times M$. By the homomorphism, $u_0 x$ is mapped on $(v_0 \times u_0)x = \Sigma_h v_0 \times (x\bar{\sigma}_h) u_h$

$= \Sigma_{h,i}((x\bar{\sigma}_h)\bar{\sigma}_i)(v_i \times u_h)$. On the other hand, $u_0x = \Sigma_h(x\bar{\sigma}_h)u_h = \Sigma_h(x\bar{\sigma}_h)u_0t_h$, and this is mapped on

$$\Sigma_h(x\bar{\sigma}_h)(v_0 \times u_0)t_h = \Sigma_h(x\bar{\sigma}_h)(v_0 \times u_h) = \Sigma_{h,i}(x\bar{\sigma}_h)(1\bar{\sigma}_i)(v_i \times u_h).$$

Hence

$$(x\bar{\sigma}_h)\bar{\sigma}_i = (x\bar{\sigma}_h)(1\bar{\sigma}_i). \tag{14}$$

So,

$$u_0(x\bar{\sigma}_h) = \Sigma_i((x\bar{\sigma}_h)\bar{\sigma}_i)u_i = \Sigma_i(x\bar{\sigma}_h)(1\bar{\sigma}_i)u_i = (x\bar{\sigma}_h)u_0.$$

Hence $x\bar{\sigma}_h \varepsilon S$ (for every $x \varepsilon A$ and every h). Also

$$\Sigma_h u_0(x\bar{\sigma}_h)t_h = \Sigma_h(x\bar{\sigma}_h)u_0t_h = \Sigma_h(x\bar{\sigma}_h)u_h = u_0x.$$

If $\mathfrak{r}$ denotes the right-ideal $\{x \varepsilon A \mid u_0x = 0\}$ of A, then

$$x - \Sigma_h(x\bar{\sigma}_h)t_h \varepsilon \mathfrak{r}. \tag{15}$$

It is clear, because of (12), that the t_h are S-left-independent modulo $\mathfrak{r}$. Hence $\{t_h\}$ forms an independent left S-basis of A mod. $\mathfrak{r}$. Suppose $\mathfrak{r}$ is 0. Then $x = \Sigma_h(x\bar{\sigma}_h)t_h$, and $\{t_h\}$ forms an independent left S-basis of A. It is then easy to verify, observing (12), (15) particularly, that M is (A-two-sided) isomorphic to $A \times {}_SA$, by the correspondence $\{u_h\} \rightarrow \{t_h\}$ over A.

Our assumption was that $u_0 \rightarrow v_0 \times u_0$ gives an (A-)homomorphism of M into $N \times M$. This, however, can be secured either by assuming $R(N \times M, v_0 \times u_0) \subseteq R(M, u_0)$, or by assuming $R^{\#}(N \times M, v_0 \times u_0) \subseteq R^{\#}(M, u_0)$, by Propositions 4, 7. This proves the second half of the following Proposition, whose first half is readily seen to be true.

PROPOSITION 10. *Let M be a special A-double-module with generator u_0. Let N be a second special A-double-module which is isomorphic to M, and v_0 its generator corresponding to u_0. M is isomorphic to $A \times {}_SA$ for some subring S of A, with u_0 corresponding to 1×1, if and only if*

$$R(N \times M, v_0 \times u_0) \subseteq R(M, u_0) \quad (\text{or } R^{\#}(N \times M, v_0 \times u_0) \subseteq R^{\#}(M, u_0))$$

and $u_0x = 0$ $(x \varepsilon A)$ implies $x = 0$.

4. Topologies in M^*, $\mathfrak{A}$. Coming back to the general case where K and A are perhaps different, we consider K, A, and M all in their discrete topologies; the particular concern is in K. We then introduce in $\mathfrak{A}$ the weak topology, in which a neighborhood of 0 is the set of elements vanishing at a finite subset of A. We consider M^* also in its weak topology. The mapping $\sigma \rightarrow \bar{\sigma}$ of M^* onto $R(M, u_0)$ is then continuous, since $x_i\bar{\sigma} = 0$ $(i = 1, 2, \cdots, m)$

are implied by $(u_0x_i)\sigma = 0$ $(i = 1, 2, \cdots, m)$. Provided that u_0 generates M, as K-A-module (i. e. $M = Ku_0A$), the mapping is open too, as we see readily; hence it is a homeomorphism, since it is 1-1 under the assumption. Further, *$R(M, u_0)$ is closed in $\mathfrak{A}$, whenever $M = Ku_0A$* (or, more generally, when every element of $(Ku_0A)^*$ can be extended to an element of M^*). To prove this, let α be an element of the closure of $R(M, u_0)$ in $\mathfrak{A}$. We then put

$$(\sum_{i=1}^{m} k_iu_0x_i)\sigma = \sum_{i=1}^{m} k_i(x_i\alpha),$$

where m is an arbitrary natural number, and k_i, x_i are arbitrary elements of K, A. This gives a unique definition of σ as a mapping of M into K. For, with different expressions of an element of M as sums of elements of the form ku_0x, the right-hand sides are equal, since the same is the case with arbitrary α in $R(M, u_0)$. σ is clearly K-linear, and $\sigma \in M^*$. Furthermore, α is equal to the $\bar{\sigma}$ given by this σ, which proves our assertion.

Assume that M possesses an independent K-basis. Then ($M^{\#}$ is dense in M^* whence) *$R^{\#}(M, u_0)$ is dense in $R(M, u_0)$*. Propositions 3, 6, combined, give

PROPOSITION 11. *Under the same assumption as in Proposition 6, $R(N \times M, v_0 \times u_0)$ is the closure of*

$$R(M, u_0)R(N, v_0) \quad (\textit{or, of } R^{\#}(M, u_0)R^{\#}(N, v_0))$$

in $\mathfrak{A}$.

5. The main theorem. Consider again the case $K = A$, and consider a special A-double-module M with generator u_0. Suppose $R(M, u_0)$ is a ring. Then Proposition 11 implies that the condition $R(N \times M, v_0 \times u_0) \subseteq R(M, u_0)$ in Proposition 10 is fulfilled. Thus we have

THEOREM. *A special A-double-module M is isomorphic to $A \times {}_SA$, with some subring S of A, if and only if the relation-module $R(M, u_0)$ forms a ring and $u_0x = 0$ implies $x = 0$.*

6. Characterization of relation-modules. So far our theory has been fairly smooth. In particular, our Theorem generalizes the similar theorem in the sfield case. Coming back to the case $K \neq A$ and seeking for a characterization of relation-modules, we encounter difficulties (cf. the examples below), which do not prevail in the sfield case. However, before we give up, let us prove

LEMMA 3. *Suppose a K-right-submodule $\mathfrak{M}$ of (the A-K-module) $\mathfrak{A}$ has an independent K-basis $\{\alpha_h\}$ such that for each $x \varepsilon A$, almost all $x\alpha_h$ are 0. Suppose further that $\mathfrak{M}$ is dense in $A \cdot \mathfrak{M}$, and in fact that for each $a \varepsilon A$ and for a finite number of h, say $1, 2, \cdots, m$, almost all $a \cdot \alpha_i$ are contained in the closure of the K-right-module spanned by $\{\alpha_j \ (j \neq 1, 2, \cdots, m)\}$. Then $\mathfrak{M}$ is the restricted relation-module of a certain K-A-double-module.*

To prove the lemma, let $\mathfrak{M}^*$ be the dual-module of the K-(right-)module $\mathfrak{M}$, i. e. the K-left-module of all continuous K-homomorphisms of $\mathfrak{M}$ into K. Since $\mathfrak{M}$ is dense in $A \cdot \mathfrak{M}$, every element of $\mathfrak{M}^*$ can be considered as a continuous K-homomorphism of $A \cdot \mathfrak{M}$ into K, and $\mathfrak{M}^*$ is essentially the dual-module of $A \cdot \mathfrak{M}$. Hence $\mathfrak{M}^*$ is a K-A-double-module. Let u_h be the elements of $\mathfrak{M}^*$ such that $u_h\alpha_i = \delta_{hi}$, and let M be the K-(left-)submodule of $\mathfrak{M}^*$ spanned by these u_h. It is the totality of elements u of $\mathfrak{M}^*$ such that almost all $u\alpha_i$ are 0 (for each u). Let u_0 be the elements of $\mathfrak{M}^*$ defined by $u_0\alpha = 1\alpha$ $(\alpha \varepsilon \mathfrak{M})$, 1 being the unit element of A. Because of our assumption, $u_0\alpha_h = 0$ for almost all h. Hence $u_0 \varepsilon M$. Further, M is A-right-allowable. To prove this, let $u \varepsilon M$, $a \varepsilon A$. Define ua by $(ua)\alpha = u(a \cdot \alpha)$. Let $\{1, 2, \cdots, m\}$ be the totality of indices h such that $u\alpha_h \neq 0$. Almost all $a \cdot \alpha_i$ are in the closure of the K-module spanned by $\{\alpha_j \ (j \neq 1, 2, \cdots, m)\}$. Hence almost all $(ua)\alpha_i$ $(= u(a \cdot \alpha_i))$ are 0, which proves $ua \varepsilon M$. It is now easy to see that $\mathfrak{M}$ can be considered as the restricted relation-module of this K-A-module M; observe

$$(u_0x)\alpha = u_0(x \cdot \alpha) = 1(x \cdot \alpha) = x\alpha \qquad (\alpha \varepsilon \mathfrak{M}). \tag{16}$$

PROPOSITION 12. *In order that a K-submodule $\mathfrak{M}$ of A is a restricted relation-module of a special K-A-double-module, it is (necessary and) sufficient that $\mathfrak{M}$ be dense in $A \cdot \mathfrak{M}$ and possess an (independent) K-basis $\{\alpha_h\}$ such that for each $a \varepsilon A$, almost all $a\alpha_h$ are 0, and there exist t_h in A satisfying $t_h\alpha_i = \delta_{hi}$.*

To prove this, we consider an arbitrary element a of A, and any finite number of h, say $1, 2, \cdots, m$. We want to show that almost all $a \cdot \alpha_i$ are in the closure of the K-module spanned by $\{\alpha_j \ (j \neq 1, 2, \cdots, m)\}$. Each $a \cdot \alpha_i$ is anyway in the closure of $\mathfrak{M}$, by assumption. As is easily seen, it is an infinite sum $\Sigma\, \alpha_h k_h^{(i)}$ $(k_h \varepsilon K)$ in the sense of convergent sum in the topology of $\mathfrak{A}$. Thus $t_h(a \cdot \alpha_i) = k_h^{(i)}$. On the other hand, $t_h(a \cdot \alpha_i) = (t_h a)\alpha_i$, and with a given h this is 0 except for a finite number of i, which proves our assertion. Thus $\mathfrak{M}$ is the restricted relation-module of the module M constructed in the proof of Lemma 3. Also $(u_0t_i)\alpha_j = t_i\alpha_j = \delta_{ij}$. Hence $u_i = u_0t_i$. It is now clear that M is a special K-A-double-module with generator u_0. (The necessity assertion in the proposition is evident.)

Further, we see readily

PROPOSITION 13. *In order that an A-K-submodule of $\mathfrak{A}$ be a relation-module of a special K-A-double-module, it is necessary and sufficient that it be closed in $\mathfrak{A}$ and possess a dense subset $\{\alpha_h\}$ such that for every $a \varepsilon A$, almost all $a\alpha_h$ are 0, and there exist elements t_h in A satisfying $t_h\alpha_i = \delta_{hi}$.*

Remark. The condition of the existence of t_h in our Propositions is very strong, and in that sense our criteria of relation-modules are very poor. If, on the other hand, K is a sfield, then the automatic existence of α_h and t_h can be proved at least for K-finite modules (and also for K-infinite modules under a suitable topological condition) ([1], [2]; cf. also § 6 below), and thus we obtain a nice theorem in that case. However, such is certainly not the case in general, and our assumption seems indispensable,[3] as the following examples show:

Example 1. Let F be an arbitrary field and K the simple algebra of all 3-dimensional matrices over F. We take A to be identical with K. Let $e_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Let α_1 be the identity mapping of K $(= A)$, and α_2 the mapping $x \to e_1x$ $(x \varepsilon K)$. It is easy to see that α_1, α_2 are K-right-independent. The module $\alpha_1K \oplus \alpha_2K (\subseteq \mathfrak{A})$ even forms a ring; the ring property is naturally of interest in connection with our theorem (in § 5). In spite of these nice properties of having an independent finite K-basis and being a ring, our module $\alpha\text{-}K \oplus \alpha_2K$ does not possess the property required in Proposition 12 (or Proposition 13),[4] as we see without difficulty; the argument is similar to that in the next example.

Example 2. We now give an example in which K is a (non-commutative) integrity-domain. Let K be the integral domain in a $\mathfrak{p}$-adic division-algebra which is of degree 2 over its center. Let π be a primitive element of $\mathfrak{p}$ such that π^2 belongs to the center. We again take $A = K$. Let α_1 be the identity mapping of K $(= A)$, and α_2 the mapping $x \to \pi x$ $(x \varepsilon K)$. Again α_1, α_2 are right-independent over K, and $\alpha_1K \oplus \alpha_2K$ is a ring. Nevertheless, this module does not fulfill the requirement of Proposition 12 (or Proposition 13). (For, if there were $\beta_1, \beta_2 \varepsilon \alpha_1K \oplus \alpha_2K$ and $t_1, t_2 \varepsilon K$ with $t_h\beta_i = \delta_{hi}$, then the matrix $(t_h\alpha_i)_{hi}$ would be regular. But $\begin{pmatrix} t_2\alpha_1 & t_2\alpha_2 \\ t_1\alpha_1 & t_1\alpha_2 \end{pmatrix} = \begin{pmatrix} t_1 & t_1\pi \\ t_2 & t_2\pi \end{pmatrix}$, and this can not possess an inverse (in the 2-dimensional matrix ring over K)).

---

[3] It is very desirable, however, to find a simpler substitute.

[4] Or, what amounts to the same, K is not a regular module of $\alpha_1K \oplus \alpha_2K$ in the sense of [4].

7. Remarks on the case where K is a sfield. If K is in particular a sfield, then naturally every K-left-module M has an independent K-basis. Moreover, if N is a K-submodule of M, then every element of N^* can be extended to an element of M^*. Hence, as Lemmas 1, 2 (in § 1, 2) show, we may restrict ourselves, without loss in generality, to K-A-double-modules M which are generated by u_0 (i. e. cyclic modules in the sense of [1], [2]). Further, such K-A-modules are all special with generator u_0 (in our sense). Thus, except for the criteria (Propositions 12, 13) of relation-modules, our results seem to offer satisfactory generalizations of corresponding theorems in the sfield case. As for Propositions 12, 13, they do not include the perfect characterization of relation-modules given in [2], Theorem 5. 3 (finite-dimensional case) and [1], Theorem 1 (infinite-dimensional case); the pathological situation in the general case being exhibited in our examples in § 5 (they show that the immediate extension of [2], Theorem 5. 3 (or [1], Theorem 1) cannot be true). It is thus necessary to clarify the relationship between Proposition 13, for instance, and [1], Theorem 1. In fact, we prove

LEMMA 4. *If K is a sfield and if an A-K-submodule of A is linearly compact (over K) then it has the property of Proposition* 13.

To show this, we have actually to borrow the argument of [1], Theorem 1. Let M be the K-A-double-module consisting of all continuous K-homomorphisms of our module into K. Let u_0 be the element of M which sends every element of our module to its value at 1 (εA). Then $M = Ku_0A$, as was shown in [1]. Let $\{u_h\}$ be an independent K-basis of M contained in u_0A, and let $u_h = u_0t_h$. Let M^* be the dual module of the K-left-module M. It is essentially our given module.[5] (The dual of the dual of a linearly compact codule is, essentially, the module itself.) In fact, if we associate with each σ in M^* the element $\bar{\sigma}$ (as defined in § 1) of $\mathfrak{A}$, then the totality of $\bar{\sigma}$'s is exactly our original submodule of $\mathfrak{A}$. Now, let σ_h and $\bar{\sigma}_h$ be as in § 3. Then $\{\sigma_h\}$ is clearly dense in M^*, and therefore $\{\bar{\sigma}_h\}$ is dense in our module. But $t_i\bar{\sigma}_k = \delta_{ik}$. That almost all $a\bar{\sigma}_h$ are 0 for each a, is clear from the fact that $a\bar{\sigma}_h$ is simply the coefficient of u_h in the K-linear expression of u_0a by $\{u_h\}$.

(This lemma explains the relationship between Proposition 13 and [1], Theorem 1. However, since the essential part of the latter is used in the proof of our lemma, we do not claim, in the least, that the latter is contained in the former, as we want to repeat. (Moreover, in our proof of the lemma

---

[5] However, we prefer not to identify the two.

we actually proved that our module is a relation-module, and after having done so, Proposition 13 is rather superfluous.) In turn, Theorem 1 of [1] depends essentially on (the properties of linearly compact spaces, such as having discrete duals and being the duals of their duals, and) [2], Lemma 2.1 (i.e. [1], Lemma 1), which is also essentially the theorem of dual bases of (finite) modules over a sfield).

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] J. Dieudonné, "Linearly compact spaces and double vector spaces over sfields," *American Journal of Mathematics*, vol. 73 (1951), pp. 13-19.

[2] G. Hochschild, "Double vector spaces over division rings," *American Journal of Mathematics*, vol. 71 (1949), pp. 443-460.

[3] N. Jacobson, "An extension of Galois theory to non-normal and non-separable fields," *American Journal of Mathematics*, vol. 69 (1947), pp. 27-36.

[4] T. Nakayama, "Non-normal Galois theory for non-commutative and non-semisimple rings," *Canadian Journal of Mathematics* (1951), pp. 208-218.

ORDER AND TOPOLOGY IN PROJECTIVE PLANES.*

By OSWALD WYLER.

1. Introduction. In a finite-dimensional projective geometry with a topological coordinate field or skew-field, the topology of the coordinate field induces in a natural way a topology of the geometry, in which the line is closed and homeomorphic to the coordinate field with an additional element at infinity. This leads to the qeustion of whether a topology of a nondesarguesian plane can be deduced in a similar manner from a topology of its lines, and whether two such planes with homeomorphic lines will be homeomorphic.

In the most general case, this seems to be impossible without complicated additional assumptions. There are, however, two important exceptions to this statement, ordered projective planes, and planes with coordinates in an alternative division ring. In both cases the answer to our question is affirmative.

The first case is discussed in this paper. No additional assumption is needed beyond the axioms of incidence and of order.

Ordered projective planes have been studied by many authors,[1] chiefly in axiomatic treatments of real projective geometry. In this paper, a topology of an ordered projective plane is deduced from the interval topology on its lines, and the main properties of this topology of the plane are investigated. The Theorem of Desargues is never needed.

In this topology, the projective operations of joining two points by a line and of intersecting two lines in a point are continuous, and two planes with homeomorphic intervals are homeomorphic. This completes the answer to the question asked above. The homeomorphism between the intervals is assumed to map endpoints on endpoints. An order-preserving mapping of an interval onto another is such a homeomorphism. We do not have to assume, however, that our homeomorphism is order-preserving.

An example of a nondesarguesian ordered plane, given in section **9**, shows that the axioms of order, even with the Dedekind continuity axiom added, still make it necessary to assume a configuration theorem in order to get coordinates.

The decomposition of an ordered plane into three convex quadrangles

* Received June 22, 1951.

[1] Cf. e.g. [1], [6], [7]. Numbers in brackets refer to the references at the end of the paper.

with the same vertices, studied in section **6**, is constantly used in the next two sections.[2] It is much more useful for our purpose than the usual decomposition of a plane into four convex triangles with the same vertices.

Points are denoted by lower case letters, lines by small gothic letters. The line joining the points a and b is denoted by ab, the point of intersection of the lines $\mathfrak{l}$ and $\mathfrak{m}$ by $\mathfrak{l} \cap \mathfrak{m}$.

2. The axioms of separation. An *ordered projective plane* is characterized by the axioms of incidence and by a relation of *separation* between pairs of points, satisfying the axioms stated below. We write $ab \parallel cd$ if two points a and b separate two points c and d. The following axioms of separation are assumed.

S. 1. If $ab \parallel cd$, then a, b, c, d are four different collinear points.

S. 2. If a, b, c, d are four different collinear points, then at least one of the relations $ab \parallel cd$, $ac \parallel bd$, $bc \parallel ad$ holds.

S. 3. If $ab \parallel cd$, then $ab \parallel dc$.

S. 4. If $ab \parallel cd$ and $bc \parallel de$, then $cd \parallel ea$.

S. 5. If $ab \parallel cd$, and if a, b, c, d are mapped on a', b', c', d' by a perspectivity, then $a'b' \parallel c'd'$.

$ab \parallel cd$ always implies $cd \parallel ab$ by S. 1 and S. 5, since $abcd$ and $cdab$ are projective for any four different collinear points. This with S. 3 shows that separation is a symmetric relation between unordered pairs of points.

At most one of the relations in S. 2 can hold at the same time, for otherwise S. 4 would lead to a contradiction to S. 1.

If a line contains only three points, $ab \parallel cd$ never holds. If we assume that there is a line with four different points, then S. 2 and S. 5 imply

S. 6. If a, b, c are three different collinear points, then there is a point d such that $ab \parallel cd$.

Applying S. 6 repeatedly together with the other axioms, we see that an ordered projective plane cannot be finite.

Our axioms of separation are essentially those of [1], p. 22. Axiom S. 4 has been put into a more suggestive form.

---

[2] Cf. [2], p. 421, where this decomposition is denoted by $\{4, 3\}/2$ (for the elliptic plane). The author is indebted to the referee for this reference.

3. Segments and intervals. If a, b, c are three different points on a line $\mathfrak{g}$, we denote by $(ab)_c$ the set of all points x with $ab \parallel cx$. A set of this form is called a *segment* on $\mathfrak{g}$. The points a and b are called the *endpoints* of the segment $(ab)_c$. The set consisting of the segment $(ab)_c$ and its endpoints a and b is denoted by $[ab]_c$. Such a set is called an *interval* on $\mathfrak{g}$. We recall some well known properties of segments.

If $ab \parallel cd$, then the segments $(ab)_c$ and $(ab)_d$ are disjoint, and every point on $\mathfrak{g}$ different from the endpoints a and b is in one of them. If t is in $(ab)_d$, then $(ab)_t = (ab)_c$. There are thus exactly two segments with given endpoints a and b on $\mathfrak{g}$. A point in one of them shall be called an *exterior* point of the other.

If p and q are in $(ab)_c$, then $(ap)_c$ and $(pq)_c$ are contained in $(ab)_c$, and b is an exterior point of $(ap)_c$ and of $(pq)_c$. The first part of this proposition remains true if we replace segments by intervals.

LEMMA 3.1. *For any point u of a segment $(ab)_c$, there exist points p and q in $(ab)_c$, such that $(pq)_c$ contains u and is contained in $(ab)_c$. Then a and b are exterior points of $(pq)_c$.*

Proof. By S. 6 there is a point p such that $ua \parallel bp$. Then $cu \parallel ab$ implies $ab \parallel pc$ and $bp \parallel cu$ by S. 4. Thus p is in $(ab)_c$, and $(bp)_c$ contains u and is continued in $(ab)_c$. Similarly there is a point q in $(bp)_c$, such that $(pq)_c$ contains u and is contained in $(bp)_c$.

LEMMA 3.2. *The intersection of two segments $(ab)_t$ and $(cd)_t$ with a common exterior point t is either empty or a segment $(pq)_t$, where p and q are two of the endpoints of the given segments.*

Proof. If u is in both segments, then c is in one of the segments $(tu)_a$ and $(tu)_b$, and d in the other. If c is in $(tu)_b$ and d in $(tu)_a$, then $c = a$ or $ct \parallel ua$ or $at \parallel uc$ by S. 2, and similarly $d = b$ or $dt \parallel ub$ or $bt \parallel ud$. If the segments are not equal we may assume $ct \parallel ua$. Then $tu \parallel ab$ implies $ab \parallel ct$ by S. 4, hence c is in $(ab)_t$. If $b = d$, then $(cd)_t$ is contained in $(ab)_t$. This holds also for $dt \parallel ub$, since then d is in $(ab)_t$. If $bt \parallel ud$, then b is in $(cd)_t$; hence $(bc)_t$ is contained in $(ab)_t$ and in $(cd)_t$, and it is easily shown to be their intersection.

THEOREM 3.3. *If two segments $\mathfrak{a}$ and $\mathfrak{b}$ on a line $\mathfrak{g}$ have a point u in common, then there exists a segment on $\mathfrak{g}$ containing u and contained in the intersection of $\mathfrak{a}$ and $\mathfrak{b}$.*

Proof. Let s be an exterior point of $\mathfrak{a}$, t an exterior point of $\mathfrak{b}$. If $s = t$, then $\mathfrak{a} \cap \mathfrak{b}$ is a segment by Lemma 3.2. If $s \neq t$, then there exists a seg-

ment $\mathfrak{c}$ containing u with s and t as exterior points by Lemma 3. 1. Then $\mathfrak{a} \cap \mathfrak{c}$ is a segment with exterior points s and t by Lemma 3. 2, and hence $\mathfrak{a} \cap \mathfrak{c} \cap \mathfrak{b}$ is a segment containing u.

Theorem 3. 3 shows that segments on a line $\mathfrak{g}$ define a topology on $\mathfrak{g}$ in which they are a base of open sets. In this topology the closure of the segment $(ab)_c$ is the interval $[ab]_c$, so that intervals form a base of closed neighborhoods. The topology is called the *interval topology* on $\mathfrak{g}$. By Lemma 3. 1, $\mathfrak{g}$ is a regular space in its interval topology.

By S. 5 a projectivity between two lines maps segments on segments and intervals on intervals, hence it is a homeomorphism between the two lines.

4. Sectors. By specifying that two pairs of lines in a pencil separate each other if and only if they intersect a line not in the pencil in pairs of points separating each other, we obtain a relation of separation for pairs of lines. Because of S. 5, this definition is independent of the choice of the auxiliary transversal line. It is immediately seen that the dual propositions to our axioms S. 1 to S. 5 hold. Hence the definitions and results of the preceding section can be dualized. The dual of a segment will be called an *angle.*

Let $\mathfrak{l}$ and $\mathfrak{m}$ be two lines with a common point r, and let a and b be two points different from r. Then we shall say that $\mathfrak{l}\mathfrak{m} \parallel ab$ whenever $\mathfrak{l}\mathfrak{m} \parallel \mathfrak{p}\mathfrak{q}$ for $\mathfrak{p} = ra$ and $\mathfrak{q} = rb$. If $\mathfrak{l}$ and $\mathfrak{m}$ are two different lines, and if c is a point neither on $\mathfrak{l}$ nor on $\mathfrak{m}$, then we denote by $(\mathfrak{l}; \mathfrak{m})_c$ the set of all points x of the plane such that $\mathfrak{l}\mathfrak{m} \parallel cx$. A set of this type is called a *sector.* The lines $\mathfrak{l}$ and $\mathfrak{m}$ are called the *sides,* and the point $\mathfrak{l} \cap \mathfrak{m}$ is called the *vertex* of the sector $(\mathfrak{l}; \mathfrak{m})_c$.

To the sector $(\mathfrak{l}; \mathfrak{m})_c$ corresponds the angle $(\mathfrak{l}\mathfrak{m})_{rc}$, where $r = \mathfrak{l} \cap \mathfrak{m}$, consisting of all lines of the pencil with vertex r through points of $(\mathfrak{l}; \mathfrak{m})_c$. Conversely, the sector $(\mathfrak{l}; \mathfrak{m})_c$ consists of all points on lines of the angle $(\mathfrak{l}\mathfrak{m})_{rc}$ and different from its vertex r, or in other words, of all points lying on exactly one line of the angle $(\mathfrak{l}\mathfrak{m})_{rc}$.

Thus the dual of a sector is the set of all lines meeting a given segment and different from the line containing the segment, or in other words, the set of all lines intersecting a given segment in exactly one point.

The following properties of sectors are immediately derived from the corresponding properties of segments on a line and of angles in a pencil.

Let $\mathfrak{l}$ and $\mathfrak{m}$ be two different lines with point of intersection r, and let c and d be two points such that $\mathfrak{l}\mathfrak{m} \parallel cd$. Then the sectors $(\mathfrak{l}; \mathfrak{m})_c$ and $(\mathfrak{l}; \mathfrak{m})_d$ are disjoint, and every point of the plane not on $\mathfrak{l}$ or on $\mathfrak{m}$ lies in one of them.

If t is in $(\mathfrak{l};\mathfrak{m})_d$, then $(\mathfrak{l};\mathfrak{m})_c = (\mathfrak{l};\mathfrak{m})_t$. There are thus exactly two sectors with given sides $\mathfrak{l}$ and $\mathfrak{m}$. A point in one of these sectors shall be called an *exterior* point of the other.

If $\mathfrak{g}$ is a line not through r, let $p = \mathfrak{g} \cap \mathfrak{l}$ and $q = \mathfrak{g} \cap \mathfrak{m}$. Then one of the segments with endpoints p and q is contained in $(\mathfrak{l};\mathfrak{m})_c$, the other in $(\mathfrak{l};\mathfrak{m})_d$. If $\mathfrak{g}$ is a line through the vertex r, different from $\mathfrak{l}$ and from $\mathfrak{m}$, then all points of $\mathfrak{g}$ different from r lie in the same sector with sides $\mathfrak{l}$ and $\mathfrak{m}$.

This implies immediately the following lemma.

LEMMA 4.1. *Let p and q be two different points of the sector $(\mathfrak{l};\mathfrak{m})_c$, and let $l = pq \cap (\mathfrak{l} \cap \mathfrak{m})c$. Then the segment $(pq)_t$ is contained in the sector $(\mathfrak{l};\mathfrak{m})_c$.*

5. The order topology of the plane. A set $\mathfrak{a}$ of points in the plane is called $\mathfrak{g}$-*convex*, where $\mathfrak{g}$ is a line, if $\mathfrak{a}$ and $\mathfrak{g}$ are disjoint, and if for any two different points p, q in $\mathfrak{a}$, the segment $(pq)_t$ is contained in $\mathfrak{a}$ for $t = pq \cap \mathfrak{g}$. A set of points is called *convex* if it is $\mathfrak{g}$-convex for some line $\mathfrak{g}$. A convex set is called *open* if its intersection with any line is an open set in the interval topology of the line.

A set consisting of one point not on $\mathfrak{g}$ is $\mathfrak{g}$-convex, and so is $\pi - \mathfrak{g}$, where π is the plane. Moreover, $\pi - \mathfrak{g}$ is open. If a and b are two different points not on $\mathfrak{g}$, and if $c = ab \cap \mathfrak{g}$, then the segment $(ab)_c$ and the interval $[ab]_c$ are $\mathfrak{g}$-convex. A sector $(\mathfrak{l};\mathfrak{m})_c$ is (rc)-convex and open for $r = \mathfrak{l} \cap \mathfrak{m}$. By Lemma 5.1 below, $(\mathfrak{l};\mathfrak{m})_c$ is also $\mathfrak{l}$-convex and $\mathfrak{m}$-convex. The intersection of any number of $\mathfrak{g}$-convex sets is again a $\mathfrak{g}$-convex set.

LEMMA 5.1. *A $\mathfrak{g}$-convex set $\mathfrak{a}$ is $\mathfrak{h}$-convex for any line $\mathfrak{h}$ disjoint with $\mathfrak{a}$.*

Proof. If p and q are different points of $\mathfrak{a}$, let $t = pq \cap \mathfrak{g}$, and let $u = pq \cap \mathfrak{h}$. Then u is an exterior point of $(pq)_t$ by assumption, hence $(pq)_u = (pq)_t$ is contained in $\mathfrak{a}$.

THEOREM 5.2. *Convex open sets define a topology of the plane of points, in which they are a base of open sets.*

Proof. Since any point of the plane is contained in some convex open set, it suffices to prove that for any two convex open sets $\mathfrak{a}$ and $\mathfrak{b}$ with a common point u, there exists a convex open set containing u and contained in $\mathfrak{a} \cap \mathfrak{b}$.

Let $\mathfrak{a}$ be $\mathfrak{l}$-convex, and let $\mathfrak{b}$ be $\mathfrak{m}$-convex, where $\mathfrak{l}$ and $\mathfrak{m}$ are two lines. If $\mathfrak{l} = \mathfrak{m}$, then $\mathfrak{a} \cap \mathfrak{b}$ is $\mathfrak{l}$-convex and also open, since the intersection of two open sets on a line is open. If $\mathfrak{l} \neq \mathfrak{m}$, let $\mathfrak{c}$ be the sector with sides $\mathfrak{l}$ and $\mathfrak{m}$

containing u. Then $\mathfrak{c}$ is $\mathfrak{l}$-convex and $\mathfrak{m}$-convex and open. Hence $\mathfrak{a} \cap \mathfrak{c}$ is $\mathfrak{l}$-convex and open, and also $\mathfrak{m}$-convex by Lemma 5.1. Thus $\mathfrak{a} \cap \mathfrak{c} \cap \mathfrak{b}$ is $\mathfrak{m}$-convex and open. It contains u and is contained in $\mathfrak{a} \cap \mathfrak{b}$.

The topology defined in Theorem 5.2 is called the *order topology* of the plane of points. The order topology of the plane of lines is defined dually. From now on, the plane of points and the plane of lines are considered as topological spaces with their order topologies. In the plane of points, lines are closed sets, and the relative topology on a line is its interval topology.

6. Convex quadrangles.

Definition. A *convex quadrangle* is the intersection of two sectors such that the vertex of each is an exterior point of the other.

If r and s are the vertices of two such sectors, then $r \neq s$, and both sectors are (rs)-convex open sets; hence so is their intersection. A convex quadrangle is thus an open convex set.

The intersection of two sectors with different vertices r and s is a convex quadrangle if and only if the line rs joining the two vertices is a common exterior line of the two angles corresponding to the sectors.

If a, b, c, d are four points, no three of which are collinear, let

$$r = ab \cap cd, \qquad s = ad \cap bc, \qquad t = ac \cap bd$$

be the diagonal points of the complete quadrangle $abcd$, and let

$$\rho_s = (ab\,;cd)_s, \qquad \sigma_r = (ad\,;bc)_r, \qquad \tau_r = (ac\,;bd)_r,$$

$$\rho_t = (ab\,;cd)_t, \qquad \sigma_t = (ad\,;bc)_t, \qquad \tau_s = (ac\,;bd)_s.$$

These notations will be used consistently.

If $\mathfrak{l} = ab$ and $\mathfrak{m} = cd$, then [3] $\mathfrak{l}\mathfrak{m} \parallel st$. Hence ρ_s and ρ_t are the two sectors with sides ab and cd, and $\rho_s \cap \sigma_r$ is a convex quadrangle containing t. It is easily seen that any convex quadrangle can be obtained in this fashion. This shows also that a convex quadrangle is never empty. The points a, b, c, d are called the *vertices*, the segments $(ab)_r$, $(bc)_s$, $(cd)_r$, $(ad)_s$, the *sides* of the quadrangle $\rho_s \cap \sigma_r$.

THEOREM 6.1. *The three convex quadrangles*

$$\rho_s \cap \sigma_r, \qquad \rho_t \cap \tau_r, \qquad \sigma_t \cap \tau_s,$$

their sides

$$(ab)_r, \quad (cd)_r, \quad (ad)_s, \quad (bc)_s, \quad (ac)_t, \quad (bd)_t,$$

and their vertices a, b, c, d, form a covering of the plane by disjoint convex sets.

---

[3] Cf. [1], section 3.21, p. 24.

Proof. Let $p = ab \cap st$. Then $ab \parallel pr$, and the segment $(ab)_r$ contains p and is contained in σ_r and in τ_r; hence it does not intersect one of the quadrangles. It is equally easily verified that all the sets of the covering are pairwise disjoint.

An exterior point of $(ab)_r$ on ab lies in $(ab)_p$, hence in $\sigma_t \cap \tau_s$. Thus a point on a side of the complete quadrangle $abcd$ always is in one of the sets of the covering.

Now let x be a point not lying on any side of the complete quadrangle $abcd$. Then x is in one of the sectors ρ_s and ρ_t, and we may assume $x \in \rho_s$. If x is in σ_r, x is in the quadrangle $\rho_s \cap \sigma_r$. Otherwise x is in σ_t. Let then $u = sx \cap ab$ and $v = sx \cap cd$. Then u is in $(ab)_r$ and hence in τ_s, and similarly v is in τ_s. Since $(uv)_s$ is the intersection of sx with ρ_s, x is in $(uv)_s$. But $(uv)_s$ is contained in the (st)-convex set τ_s, and so x is in the quadrangle $\sigma_t \cap \tau_s$. This completes the proof.

LEMMA 6.2. *The boundary of a convex quadrangle consists of its sides and vertices.*

Proof. The complement of the open set $\rho_t \cup \sigma_t$ consists of the quadrangle $\rho_s \cap \sigma_r$ with its sides and vertices, hence this is a closed set. If u is on a side of $\rho_s \cap \sigma_r$ or one of its vertices, let $v = ut \cap rs$. Then the segment $(ut)_v$ is contained in $\rho_s \cap \sigma_r$, and u is a boundary point of $(ut)_v$, hence also of $\rho_s \cap \sigma_r$.

LEMMA 6.3. *If p is in the segment $(ab)_r$, and q in $(cd)_r$, then one of the two segments with endpoints p and q lies in $\rho_s \cap \sigma_r$, and the other in $\rho_t \cap \tau_r$.*

Proof. Let $u = pq \cap rs$ and $v = pq \cap rt$. Then $pq \parallel uv$, and $(pq)_u$ is contained in $\rho_s \cap \sigma_r$, $(pq)_v$ in $\rho_t \cap \tau_r$.

If the dual of a convex quadrangle is called a *convex quadrilateral*, then the lines intersecting two segments on two different lines form a convex quadrilateral if and only if the point of intersection of the lines is a common exterior point for the segments. Thus the lines joining a point of $(ab)_r$ and a point of $(cd)_r$ form a convex quadrilateral. By Lemma 6.3, the lines of this quadrilateral do not meet the quadrangle $\sigma_t \cap \tau_s$ or its boundary. It is easily seen that every convex quadrilateral can be obtained in this fashion.

7. Continuity theorems.

LEMMA 7.1. *Let x, y, z be three points not on a line, and let $\mathfrak{a}$ be an open convex set containing x. Then there is a complete quadrangle $abcd$ with diagonal points $r = y$ and $s = z$ such that the convex quadrangle $\rho_s \cap \sigma_r$ contains x and is contained in $\mathfrak{a}$ together with its boundary.*

Proof. If $\mathfrak{a}$ is $\mathfrak{l}$-convex for a line $\mathfrak{l} \neq yz$, let $\mathfrak{b}$ be the sector with sides $\mathfrak{l}$ and yz containing x. Then $\mathfrak{a} \cap \mathfrak{b}$ is (yz)-convex and open. Hence we may assume $\mathfrak{a}$ to be (yz)-convex. Then there is a segment $(uv)_z$ on xz containing x and contained in $\mathfrak{a}$ together with its endpoints. Since the projection of uy on vy from z maps segments on segments, there are segments $(ab)_y$ on uy and $(cd)_y$ on vy, contained in $\mathfrak{a}$ together with their endpoints, such that $ab \parallel uy$ and $ad \cap bc = z$. Then $cd \parallel vy$, and $abcd$ is the desired quadrangle. For $r = y$, $s = z$, and x in $(uv)_s$ lies in $\rho_s \cap \sigma_r$. The vertices of $\rho_s \cap \sigma_r$ lie in the (rs)-convex set $\mathfrak{a}$; hence the quadrangle and its sides are contained in $\mathfrak{a}$.

Lemma 7. 1 shows that the plane of points is a regular topological space, and that convex quadrangles form a base of open sets.

THEOREM 7. 2. *The point of intersection of two different lines* $\mathfrak{l}$ *and* $\mathfrak{m}$ *is a continuous function of the pair* $(\mathfrak{l}, \mathfrak{m})$.

Proof. By Lemma 7. 1 we may assume that a neighborhood of the point $\mathfrak{l} \cap \mathfrak{m}$ is a convex quadrangle $\rho_s \cap \sigma_r$ with r on $\mathfrak{m}$ and s on $\mathfrak{l}$. Then the lines intersectiong $(ab)_r$ and $(cd)_r$ form a neighborhood L of $\mathfrak{l}$, and the lines intersecting $(ad)_s$ and $(bc)_s$ form a neighborhood M of $\mathfrak{m}$. A line in L does not meet $\sigma_t \cap \tau_s$ or its boundary by Lemma 6. 3, and a line in M does not meet $\rho_t \cap \tau_r$ or its boundary. Then it follows from Theorem 6. 1 that the point of intersection of a line in L and a line in M must be in $\rho_s \cap \sigma_r$, and this proves the theorem.

THEOREM 7. 3. *The line joining two different points* a *and* b *is a continuous function of the pair* (a, b).

This is the dual of Theorem 7. 2.

8. Homeomorphism theorems. We shall use closed neighborhoods in this section rather than open neighborhoods. The closure of a convex quadrangle, consisting of the quadrangle with its sides and vertices, shall be called a *projective square.* The closure of $\rho_s \cap \sigma_r$ will be denoted by T, the closure of $\rho_t \cap \tau_r$ by S, and the closure of $\sigma_t \cap \tau_s$ by R. The *sides* of T are the intervals $[ab]_r$, $[bc]_s$, $[cd]_r$, and $[ad]_s$.

Any two intervals in the plane are homeomorphic, since projectivities are homeomorphisms. If the theorem of Desargues is valid in an ordered plane, then all intervals are homeomorphic to the unit interval in the ordered coordinate field or skew-field. If we speak of a homeomorphism between two intervals, we always assume that endpoints are mapped on endpoints.

THEOREM 8. 1. *A projective square is homeomorphic to the cartesian product of two intervals.*

Proof. For any point u in T, let $x = ab \cap us$ and $y = ad \cap ur$. Then x in $[ab]_r$ and y in $[ad]_s$ are continuous functions of u, and $u = xs \cap yr$. Thus u is also a continuous function of the pair (x, y), so that this correspondence is a homeomorphism between T and the product $[ab]_r \times [ad]_s$.

If $abcd$ and $a'b'c'd'$ are complete quadrangles in two ordered projective planes with homeomorphic intervals, then there is a homeomorphism between the boundaries of the projective squares T and T', by which the vertices and sides of T are mapped on the corresponding vertices and sides of T'. Such a homeomorphism shall be called a *p-homeomorphism,* where p means " proper." A homeomorphism between T and T' will be called a *P-homeomorphism* if it induces a p-homeomorphism between the boundaries.

LEMMA 8. 2. *There is a P-homeomorphism between T and T'.*

Proof. There are homeomorphisms ϕ between $[ab]_r$ and $[a'b']_{r'}$ and ψ between $[ad]_s$ and $[a'd']_{s'}$ such that $\phi(a) = \psi(a) = a'$. For any point u of T, let $x' = \phi(ab \cap us)$, $y' = \psi(ad \cap ur)$, and $\Phi(u) = x's' \cap y'r'$. Then Φ is a P-homeomorphism.

LEMMA 8. 3. *For every p-homeomorphism ϕ between the boundaries of T and of T', there is a P-homeomorphism Φ between T and T' that agrees with ϕ on the boundary.*

Proof. Since the product of two P-homeomorphisms is a P-homeomorphism, it suffices by Lemma 8. 2 to prove Lemma 8. 3 for $T = T'$. Again, it is sufficient to consider p-homeomorphisms which leave three sides of T fixed, since every p-homeomorphism on the boundary of T is a product of p-homeomorphisms of this special type.

Now let ϕ be a homeomorphism of $[ab]_r$ onto itself with $\phi(a) = a$ and $\phi(b) = b$. Let A be the closure of the open set $\tau_r \cap (ab;tr)_s$. Then A is contained in T, and its intersection with the boundary of T is the interval $[ab]_r$. The boundary of A consists of the three intervals $[ab]_r$, $[at]_c$, and $[bt]_d$. For u in A, $u \neq t$, the point $x = ab \cap tu$ is in $[ab]_r$. Now define $\Phi(u) = \phi(x)t \cap ru$ for u in A, $u \neq t$, and define $\Phi(u) = u$ for u in $T - A$ or $u = t$. Then $\Phi(x) = \phi(x)$ for x in $[ab]_r$, and $\Phi(u) = u$ for u in $[at]_c$ or in $[bt]_d$. It follows that Φ is continuous everywhere in T, except possibly at t.

If a'' is a point of $[at]_c$, then $b'' = bt \cap ra''$ is in $[bt]_c$, and b'',

$d'' = dt \cap sa''$, and $c'' = b''s \cap d''r$ are continuous functions of a''. If $a'' = t$, then $b'' = c'' = d'' = t$. Now let $\mathfrak{a}$ be a convex neighborhood of t. We can choose a point a'' in $[at]_c$, different from t, such that a'', b'', c'', and d'' are in $\mathfrak{a}$. Then the projective square T'' with these vertices is a closed neighborhood of t contained in $\mathfrak{a}$. It follows from the construction that $\Phi(T'') = T''$, hence Φ is continuous at t.

Since the inverse mapping of Φ can be constructed in the same way as Φ, Φ is a P-homeomorphism of T onto itself which agrees with ϕ on $[ab]_r$ and leaves the other three sides of T fixed. This proves the lemma.

THEOREM 8.4. *Two ordered projective planes with homeomorphic intervals are homeomorphic.*

Proof. Let $abcd$ be a complete quadrangle in one plane, $a'b'c'd'$ a complete quadrangle in the other one. Then there is a mapping ϕ of the sides and vertices of R, S, and T on the sides and vertices of R', S', and T', which determines a p-homeomorphism for each of the three pairs of corresponding projective squares. Then by Lemma 8.3 there are homeomorphisms between R and R', between S and S', and between T and T' which agree with ϕ on the boundaries. Since two squares in the same plane have no common interior point, these homeomorphisms define a homeomorphism between the two planes.

COROLLARY 1. *In an ordered projective plane, the plane of lines is homeomorphic to the plane of points.*

COROLLARY 2. *If an interval in an ordered projective plane π is homeomorphic to the unit interval in an ordered field or skew-field K, then π is homeomorphic to the coordinate plane over the field K.*

It should be remarked that the homeomorphism of Theorem 8.4 need not be a collineation. In fact, section **9** gives an example of two ordered projective planes which are homeomorphic, but not collinear.

9. A nondesarguesian ordered plane. Let K be any ordered field or skew-field. We define points by homogeneous coordinates (x, y, z) over K with right multiplication, lines by homogeneous coordinates (a, b, c) with left multiplication.

A point (x, y, z) shall be on a line (a, b, c) with $ab \geqq 0$ if and only if

$$ax + by + cz = 0.$$

For a line (a, b, c) with $ab < 0$, we distinguish three cases. A point (x, y, z) shall be on this line if and only if

$$ax + by + cz = 0 \text{ for } xz \leqq 0,$$

$$\tfrac{1}{2}ax + by + cz = 0 \text{ for } x \text{ between } 0 \text{ and } z,$$

$$ax + by + (c - \tfrac{1}{2}a)z = 0 \text{ for } z \text{ between } 0 \text{ and } x.$$

This is a modification of the well known example of a nondesarguesian affine plane given by Hilbert and originally by F. R. Moulton.[4]. The verification of the axioms of incidence is quite easy, but rather lengthy. If we define a separation relation for pairs of points in the obvious manner, then the axioms S. 1 to S. 5 are readily verified.

Consider now the triangles abc and $a'b'c'$ with vertices

$$a: (0, 1, 2); \quad b: (1, 1, 2); \quad c: (1, 1, 1);$$

$$a': (0, -1, 2); \quad b': (1, -1, 2); \quad c': (1, -1, 1).$$

These triangles are perspective with center $(0, 1, 0)$, but the points

$$ab \cap a'b': (1, 0, 0); \quad bc \cap b'c': (0, 0, 1); \quad ac \cap a'c': (4, 1, -6)$$

are not collinear.

By Theorem 8. 4, Corollary 2, our plane is homeomorphic to the coordinate plane over the field K, but the two planes cannot be collinear.

NORTHWESTERN UNIVERSITY.

REFERENCES.

[1] H. S. M. Coxeter, *The real projective plane*, New York, 1949.

[2] ———, "Self-dual configurations and regular graphs," *Bulletin of the American Mathematical Society*, vol. 56 (1950), pp. 413-455.

[3] Marshall Hall, "Projective planes," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 229-277.

[4] David Hilbert, *Grundlagen der Geometrie*, 7. Aufl., Leipzig und Berlin, 1930.

[5] F. R. Moulton, "A simple non-desarguesian plane geometry," *Transactions of the American Mathematical Society*, vol. 3 (1902), pp. 192-195.

[6] Ernst Steinitz, *Vorlesungen über die Theorie der Polyeder*, Berlin, 1934.

[7] Oswald Veblen and J. W. Young, *Projective geometry*, Boston, vol. I, 1910, vol. II, 1916.

[4] [4], p. 85, and [5].

MEANS IN GROUPS.*

By W. R. Scott.

1. Introduction. A number of authors have given sets of postulates for the arithmetic mean of n real numbers. Several of these sets of postulates have been given in purely algebraic form (see [2], [3], [4], [5], [7], [8]), and some of these latter are suitable for generalization to groups. Only the set due to Schimmack [7] will be discussed.

Let G be a (not necessarily Abelian) group written additively. Let $x_i \varepsilon G, i = 1, \cdots, n$, and let $f_n(x_1, \cdots, x_n) \varepsilon G$. Schimmack's postulates are:

(1) $\quad f_n(h + x_1, \cdots, h + x_n) = h + f_n(x_1, \cdots, x_n)$ for all $h \varepsilon G$.

(2) $\quad f_n(-x_1, \cdots, -x_n) = -f_n(x_1, \cdots, x_n)$.

(3) $\quad f_n$ is a symmetric function of $x_1, \cdots, x_n$.

(4) $\quad f_{n+1}(f_n, \cdots, f_n, x_{n+1}) = f_{n+1}(x_1, \cdots, x_{n+1})$.

In (4) and throughout the paper f_n will be used for $f_n(x_1, \cdots, x_n)$ whenever no confusion will result. A sequence $\{f_n\}$, $n = 1, 2, \cdots$, of functions will be called a *sequence* for brevity. A sequence satisfying (1), (2), (3), (4) will be called a *mean* on G.

Schimmack [7] showed that if G is the additive group R of reals, then the only mean is the ordinary arithmetic mean $f_n = (x_1 + \cdots + x_n)/n$. Beetle [1] showed that if $G = R$, then the postulates (1), (2), (3), (4) are completely independent. It will be shown here that, more generally, if G is an infinite Abelian torsion-free group such that $mG = G$ for all positive integers m, then Schimmack's and Beetle's conclusions hold. The question of existence and uniqueness of means, together with the complete independence of (1), (2), (3), (4), will be treated as completely as we are able.

2. Existence of means. If $\{f_n\}$ satisfies (1) and (2), then it follows readily that for all $h \varepsilon G$,

$$(1') \qquad f_n(x_1 + h, \cdots, x_n + h) = f_n(x_1, \cdots, x_n) + h.$$

If $\{f_n\}$ satisfies (4), then by induction, for $1 \leqq r \leqq n$,

$$(4') \qquad f_{n+1}(f_r, \cdots, f_r, x_{r+1}, \cdots, x_{n+1}) = f_{n+1}(x_1, \cdots, x_{n+1}).$$

* Received September 5, 1951.

THEOREM 1. *A group G possesses a mean if and only if*

(i) *the equation $mx = g$ possesses a unique solution (denoted by $x = g/m$) for every $g \varepsilon G$ and every positive integer m; and*

(ii) $y + (-y + nz)/(n+1) = z + (-z + ny)/(n+1)$ *for every* $y \varepsilon G$, $z \varepsilon G$, $n \geqq 2$.

If a mean exists, it is unique and is given by

$$\text{(5)} \qquad \begin{cases} f_1(x_1) = x_1, \\ f_n = f_{n-1} + (-f_{n-1} + x_n)/n. \end{cases}$$

Proof. Parts of the argument are the same as in [7] but they will be given here for the sake of completeness.

Suppose first that a mean $\{f_n\}$ exists.

G is torsion-free. For if $g \neq 0$ is of finite order m,

then by (1) and (3) we have

$$\begin{aligned} f_m(g, 2g, \cdots, (m-1)g, 0) &= g + f_m(0, g, \cdots, (m-1)g) \\ &= g + f_m(g, 2g, \cdots, (m-1)g, 0), \end{aligned}$$

which is a contradiction.

By (2), $f_n(0, \cdots, 0) = -f_n(0, \cdots, 0)$, whence $f_n(0, \cdots, 0) = 0$, since G is torsion-free. Hence $f_n(g, \cdots, g) = g$ by (1). Thus in particular $f_1(x_1) = x_1$, as asserted in (5).

We assert that

$$\text{(6)} \qquad nf_n(g, 0, \cdots, 0) = g.$$

In fact, $f_2(g, 0) = g + f_2(0, -g) = g - f_2(0, g) = g - f_2(g, 0)$, and (6) is true for $n = 2$. Assume that (6) holds for $n = 2^s$. Let $f_{2^s}(g, 0, \cdots, 0) = y$. Thus $2^s y = g$. Let z be such that $2z = y$ (such a z exists by (6) with $n = 2$). Thus $2^{s+1}z = g$. Then by (4′)

$$\begin{aligned} &2^{s+1} f_{2^{s+1}}(g, 0, \cdots, 0) \\ &\quad = 2^{s+1} f_{2^{s+1}}\ (f_{2^s}(g, 0, \cdots, 0), \cdots, f_{2^s}(g, 0, \cdots, 0), 0, \cdots, 0) \\ &\quad = 2^{s+1} f_{2^{s+1}}(y, \cdots, y, 0, \cdots, 0) \\ &\quad = 2^{s+1}(z + f_{2^{s+1}}(z, \cdots, z, -z, \cdots, -z) = g, \end{aligned}$$

the last equality following from the fact that

$$f_{2^{s+1}}(z, \cdots, z, -z, \cdots, -z) = -f_{2^{s+1}}(z, \cdots, z, -z, \cdots, -z),$$

whence $f_{2^s+1}(z, \cdots, z, -z, \cdots, -z) = 0$. Thus by induction (6) is true for $n = 2^r$, $r = 0, 1, 2, \cdots$. Suppose (6) true for $n+1$. Let $y = f_n(g, 0, \cdots, 0)$. Then

$$\begin{aligned} g &= (n+1)f_{n+1}(g, 0, \cdots, 0) \\ &= (n+1)f_{n+1}(f_n(g, 0, \cdots, 0), \cdots, f_n(g, 0, \cdots, 0), 0) \\ &= (n+1)f_{n+1}(y, \cdots, y, 0) = (n+1)(y + f_{n+1}(0, \cdots, 0, -y)) \\ &= (n+1)y - y = ny, \end{aligned}$$

the next to the last equality holding since y and $f_{n+1}(0, \cdots, 0, -y)$ commute, because $-y = (n+1)f_{n+1}(0, \cdots, 0, -y)$. Hence (6) is true by induction.

It follows from (6) that the equation $mx = g$ has at least one root. For $m = 1$ it has exactly one root. Again $f_2(2g, 0) = g + f_2(g, -g) = g$, and (i) is true for $m = 2$. Assume that (i) is true for $m = 1, \cdots, n-1$, where $n \geqq 3$. Then by (1)-(4), (6),

$$\begin{aligned} &f_n(ny, 0, \cdots, 0) \\ &\quad = f_n(f_{n-1}(ny, 0, \cdots, 0), \cdots, f_{n-1}(ny, 0, \cdots, 0), 0) \\ &\quad = f_n(ny/(n-1), \cdots, ny/(n-1), 0) \\ &\quad = f_n(f_{n-1}(y, (n-1)y/(n-2), \cdots, (n-1)y/(n-2)), \\ &\qquad\qquad \cdots, f_{n-1}(y, (n-1)y/(n-2), \cdots, (n-1)y/(n-2)), 0) \\ &\quad = f_n(y, (n-1)y/(n-2), \cdots, (n-1)y/(n-2), 0) \\ &\quad = f_n(f_{n-1}((n-1)y/(n-2), \cdots, (n-1)y/(n-2), 0), \\ &\qquad\qquad \cdots, f_{n-1}((n-1)y/(n-2), \cdots, (n-1)y/(n-2), 0), y) \\ &\quad = f_n(y, \cdots, y) = y. \end{aligned}$$

Thus $ny = nz$ implies

$$y = f_n(ny, 0, \cdots, 0) = f_n(nz, 0, \cdots, 0) = z,$$

and (i) has been proved by induction.

For $n > 1$,

$$\begin{aligned} f_n &= f_n(f_{n-1}, \cdots, f_{n-1}, x_n) \\ &= f_{n-1} + f_n(0, \cdots, 0, -f_{n-1} + x_n) = f_{n-1} + (-f_{n-1} + x_n)/n, \end{aligned}$$

and (5) has been verified. The uniqueness of the mean follows from (i) and (5).

By (3), we have

$$f_{n+1}(x_1, \cdots, x_{n+1}) = f_{n+1}(x_1, \cdots, x_{n-1}, x_{n+1}, x_n), \tag{7}$$

whence by (5),

$$(8)\quad f_{n-1} + (-f_{n-1} + x_n)/n + ((-x_n + f_{n-1})/n - f_{n-1} + x_{n+1})/(n+1)$$
$$= f_{n-1} + (-f_{n-1} + x_{n+1})/n + ((-x_{n+1} + f_{n-1})/n - f_{n-1} + x_n)/(n+1).$$

Let y and z be given. Choose any $x_1, \cdots, x_{n-1}$, and then choose x_n and x_{n+1} so that

$$(9)\quad \begin{cases} ny = -f_{n-1} + x_{n+1} \\ nz = -f_{n-1} + x_n. \end{cases}$$

Making the substitutions (9) in equation (8), we get (ii).

Conversely suppose that (i) and (ii) are satisfied by G. Define $\{f_n\}$ by (5). Evidently $f_1(x_1) = x_1$ satisfies (1), (2), (3), (and (4) vacuously) for $n = 1$.

Assume that (1) holds for a certain n. Then

$$f_{n+1}(h + x_1, \cdots, h + x_{n+1})$$
$$= f_n(h + x_1, \cdots, h + x_n) + (-f_n(h + x_1, \cdots, h + x_n) + h + x_{n+1})/(n+1)$$
$$= h + f_n + (-f_n - h + h + x_{n+1})/(n+1) = h + f_{n+1},$$

and (1) holds for all n.

Assume that (2) holds for given n. Then

$$f_{n+1}(-x_1, \cdots, -x_{n+1})$$
$$= f_n(-x_1, \cdots, -x_n) + (-f_n(-x_1, \cdots, -x_n) - x_{n+1})/(n+1)$$
$$= -f_n + (f_n - x_{n+1})/(n+1).$$

Now $(n+1)(-f_n + (f_n - x_{n+1})/(n+1) + f_n) = -x_{n+1} + f_n$. Hence

$$-f_n + (f_n - x_{n+1})/(n+1) = (-x_{n+1} + f_n)/(n+1) - f_n.$$

Thus

$$f_{n+1}(-x_1, \cdots, -x_{n+1}) = (-x_{n+1} + f_n)/(n+1) - f_n$$
$$= -(f_n + (-f_n + x_{n+1})/(n+1)) = -f_{n+1}.$$

Hence (2) holds for all n.

Re (4), it follows by induction that $f_n(x, \cdots, x) = x$. Therefore

$$f_{n+1}(f_n, \cdots, f_n, x_{n+1})$$
$$= f_n(f_n, \cdots, f_n) + (-f_n(f_n, \cdots, f_n) + x_{n+1})/(n+1)$$
$$= f_n + (-f_n + x_{n+1})/(n+1) = f_{n+1}.$$

Hence (4) holds for all n.

As noted above, (3) holds for $n = 1$. Now $-x_2 + x_1 + \frac{1}{2}(-x_1 + x_2) = \frac{1}{2}(-x_2 + x_1)$. Hence $x_1 + \frac{1}{2}(-x_1 + x_2) = x_2 + \frac{1}{2}(-x_2 + x_1)$, or $f_2(x_1, x_2) = f_2(x_2, x_1)$, and (3) holds for $n = 2$. Assume that (3) holds for a given $n \geqq 2$. Then to prove (3) inductively it is sufficient by (4) to prove (7), i. e. (8). But (8) follows from (ii) by means of the substitution (9). Thus (3) is proved by induction, and (5) defines a mean on G.

COROLLARY. *If G is an Abelian torsion-free group such that $mG = G$ for all integers $m > 0$, then G has a unique mean given by*

$$f_n(x_1, \cdots, x_n) = (x_1 + \cdots + x_n)/n.$$

Proof. It is easily verified that G satisfies conditions (i) and (ii) of Theorem 1. Equation (5) yields the above mean.

There remains the question of the existence of a non-Abelian group satisfying conditions (i) and (ii) of Theorem 1.

3. Complete independence. In order to discuss complete independence of the postulates (1)-(4) it is necessary to discuss the existence of 16 types of sequences. For brevity these types will be numbered as follows:

1.	$++++$	5.	$+-++$	9.	$-+++$	13.	$--++$
2.	$+++-$	6.	$+-+-$	10.	$-++-$	14.	$--+-$
3.	$++-+$	7.	$+--+$	11.	$-+-+$	15.	$---+$
4.	$++--$	8.	$+---$	12.	$-+--$	16.	$----$,

where a sequence is of type 7, for example, if (1) and (4) are true while (2) and (3) are false. The history (i. e. as to existence or non-existence) of each type of sequence will be discussed as completely as we are able. This has already been done for type 1. For the group G of order 1, clearly sequences of types 2 through 16 do not exist. *From now on assume that G has at least 2 elements.* Let $g \varepsilon G$, $g \neq 0$, be a fixed element.

LEMMA 1. *Every group G possesses sequences of types* 3, 4, 9, 10, 11, 12.

Proof. They will be exhibited.[1]

3. Let $f_n = x_1$.

4. Let $f_n = x_1$ if $n \neq 2$, and let $f_2 = x_2$. Re (4),

$$f_3(g, 0, 0) = g \neq 0 = f_3(f_2(g, 0), f_2(g, 0), 0).$$

[1] Several of the examples used in the proofs of Lemmas 1 and 2 were obtained by L. A. Colquitt and the author while working on the real case.

9. Let $f_n = 0$.

10. Let $f_2(0, x) = f_2(x, 0) = x$ for all $x \varepsilon G$, and let $f_n = 0$ otherwise. Concerning (4), we have $f_2(g, 0) = g \neq 0 = f_2(f_1(g), 0)$.

11. Let $f_1 = 0$ and $f_n = x_n$ for $n > 1$.

12. Let $f_1 = 0$ and $f_n = x_1$ for $n > 1$.

LEMMA 2. *A group G possesses sequences of types 7, 8, 13, 14, 15, 16 if and only if G has an element y of order greater than 2.*

Proof. If G has no element of order greater than 2, then $-x = x$ for all $x \varepsilon G$, and

$$f_n(-x_1, \cdots, -x_n) = f_n(x_1, \cdots, x_n) = -f_n(x_1, \cdots, x_n).$$

Therefore (2) is satisfied, and the listed types (and 5 and 6) do not exist.

Conversely suppose that G has an element y of order greater than 2. Then the following sequences will show the required existence.

7. Let $f_n = x_n + y$.

8. Let $f_n = x_1 + y$.

13. Let $f_n = y$.

14. Let $f_2(0, x) = f_2(x, 0) = x$ for all $x \varepsilon G$, and let $f_n = y$ otherwise. Re (4), we have $f_2(0, 0) = 0 \neq y = f_2(f_1(0), 0)$.

15. Let $f_n = -x_n + y$.

16. Let $f_n = -x_1 + y$.

LEMMA 3. *A sequence of type 6 exists if and only if G is torsion-free.*

Proof. Suppose that a sequence of type 6 exists. Then the first part of the proof of Theorem 1 shows that G must be torsion-free.

Conversely suppose that G is torsion-free. Let

$$(x_1, \cdots, x_n) \underset{h}{\sim} (y_1, \cdots, y_n)$$

if $x_r = h + y_{i_r}$, $r = 1, \cdots, n$, for some permutation $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$. If also $(x_1, \cdots, x_n) \underset{h'}{\sim} (y_1, \cdots, y_n)$, then $h + y_{j_1} = h' + y_{j_2}$, $h + y_{j_2} = h' + y_{j_3}, \cdots, h + y_{j_m} = h' + y_{j_1}$. Hence $y_{j_1} = m(-h + h') + y_{j_1}$, and $m(-h + h') = 0$. But since G is torsion-free, $-h + h' = 0$, and $h = h'$.

Let $(x_1, \cdots, x_n) \sim (y_1, \cdots, y_n)$ if $(x_1, \cdots, x_n) \underset{h}{\sim} (y_1, \cdots, y_n)$ for

some (unique) h. Clearly $\sim$ is an equivalence relation. Choose a representative $(x_1^*, \cdots, x_n^*)$ in each equivalence class $\{(x_1, \cdots, x_n)\}$, choosing (g), $(g, 0)$ and $(g, -g)$ as the representatives of their respective equivalence classes, where g is a fixed element $\neq 0$. Note that $\{(g,0)\} \neq \{(g, -g)\}$. Let $(x_1, \cdots, x_n)$ be given. Then $(x_1, \cdots, x_n) \underset{h}{\sim} (x_1^*, \cdots, x_n^*)$. Define $f_n(x_1, \cdots, x_n) = h$. Now (1) and (3) are obviously satisfied. Re (2), $f_1(-0) = -g \neq g = -f_1(0)$. Re (4),

$$f_2(2g, 0) = g \neq 0 = f_2(g, 0) = f_2(f_1(2g), 0).$$

The only types not yet discussed are types 2 and 5. The results on these are only partial.

Lemma 4. *If G possesses a sequence of type 2, then G is torsion free and possesses a unique solution x of the equation $2x = g$ for all $g \in G$.*

Proof. This was shown in the proof of Theorem 1, where (4) was not used to prove the existence and uniqueness of such solutions.

Lemma 5. *Let G be such that*

(i) *$2x = g$ always has a solution for x;*

(ii) *$nx = g$, $n > 1$, has at most one solution for x for any $g \in G$. Then a sequence of type 2 exists.*

Proof. Because of its length, only an outline of the proof will be given.

Let $(x_1, \cdots, x_n) R_1 (y_1, \cdots, y_n)$ if there exist $h, k \in G$ and a permutation $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$ such that

$$x_r = h + y_{i_r} + k, \qquad r = 1, \cdots, n.$$

Let $(x_1, \cdots, x_n) R_2 (y_1, \cdots, y_n)$ if there exist $h, k \in G$ and a permutation $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$ such that

$$x_r = h - y_{i_r} + k, \qquad r = 1, \cdots, n.$$

Let $(x_1, \cdots, x_n) \sim (y_1, \cdots, y_n)$ if either $(x_1, \cdots, x_n) R_1 (y_1, \cdots, y_n)$ or $(x_1, \cdots, x_n) R_2 (y_1, \cdots, y_n)$. It follows that $\sim$ is an equivalence relation. In each equivalence class choose a representative $(x_1^*, \cdots, x_n^*)$, letting $(0, g)$, $(0, g, 3g)$, and $(\frac{1}{2}g, \frac{1}{2}g, 3g)$ be the representatives of their (distinct) equivalence classes, where g is a fixed element $\neq 0$.

Define f_n as follows:

$$f_n(x_1^*, \cdots, x_n^*) = \begin{cases} x_1^* + \frac{1}{2}(-x_1^* + x_{i_1}^*) \\ \qquad \text{if } (x_1^*, \cdots, x_n^*) R_2 (x_1^*, \cdots, x_n^*), \\ x_1^* \text{ otherwise}; \end{cases}$$

$$f_n(h + x_{j_1}^* + k, \cdots, h + x_{j_n}^* + k) = h + f_n(x_1^*, \cdots, x_n^*) + k,$$

$$f_n(h - x_{j_1}^* + k, \cdots, h - x_{j_n}^* + k) = h - f_n(x_1^*, \cdots, x_n^*) + k.$$

It can then be shown that f_n is well defined, and satisfies (1), (2) and (3). Re (4), we have

$$f_3(f_2(0, g), f_2(0, g), 3g) = f_3(\tfrac{1}{2}g, \tfrac{1}{2}g, 3g) = \tfrac{1}{2}g \neq 0 = f_3(0, g, 3g).$$

LEMMA 6. *If G can be ordered so that $x < y$ implies $h + x < h + y$ for all $h \varepsilon G$, then a sequence of type 5 exists.*

Proof. Define $f_n(x_1, \cdots, x_n) = \min(x_1, \cdots, x_n)$. Then (1), (3) and (4) are satisfied. Now $f_2(g, -g) = f_2(-g, g) = g$ or $-g$. Since $g \neq -g$ (such groups G are torsion-free), (2) is not satisfied.

See [6] for a discussion of such groups with the additional restriction that $x < y$ imply $x + h < y + h$ for all $h \varepsilon G$.

THEOREM 2. *If G is an Abelian torsion-free group such that $nG = G$ for all integers $n > 0$, then postulates (1), (2), (3), (4) are completely independent.*

Proof. Such a group G can be ordered (see [6]). Hence by Lemma 6 a sequence of type 5 exists. The other types exist by the Corollary to Theorem 1 and Lemas 1, 2, 3 and 5.

Sequences of types 2 and 6 can be replaced by simpler sequences in this case by making use of an ordering of G.

2. Let $f_n(x_1, \cdots, x_n) = \frac{1}{2}(\min + \max)$.

6. Let $f_n = \min + g$, where g is a fixed element $\neq 0$.

THEOREM 3. *If G is the additive group of reals, then there exists a unique mean on G, namely $f_n = (x_1 + \cdots + x_n)/n$, and (1), (2), (3), (4) are completely independent.*

This follows from Theorem 2 and the Corollary to Theorem 1. A similar theorem for the geometric mean of positive real numbers can be given, of course.

UNIVERSITY OF KANSAS.

REFERENCES.

[1] R. D. Beetle, "On the complete independence of Schimmack's postulates for the arithmetic mean," *Mathematische Annalen*, vol. 76 (1915), pp. 444-446.

[2] ———, *Bulletin of the American Mathematical Society*, vol. 22 (1916), pp. 276-277.

[3] E. L. Dodd, "The complete independence of certain properties of means," *Annals of Mathematics*, vol. 35 (1934), pp. 740-747.

[4] E. V. Huntington, "Sets of independent postulates for the arithmetic mean, the geometric mean, the harmonic mean, and the root mean square," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 1-22.

[5] S. Narumi, "Note on the law of the arithmetical mean," *Tôhoku Mathematical Journal*, vol. 30 (1929), pp. 19-21.

[6] B. H. Neumann, "On ordered groups," *American Journal of Mathematics*, vol. 71 (1949), pp. 1-18.

[7] R. Schimmack, "Der Satz vom arithmetischen Mittel in axiomatischer Begründung," *Mathematische Annalen*, vol. 68 (1909), pp. 125-132.

[8] O. Suto, "Law of the arithmetical mean," *Tôhoku Mathematical Journal*, vol. 6 (1914), pp. 79-81.

A PROOF OF THE MAXIMAL CHAIN THEOREM.*

By Orrin Frink.

The maximal chain theorem was first proved by Hausdorff in 1914 in [3] using transfinite induction. It states that every chain in a partially ordered set is contained in a maximal chain. It is equivalent to Zorn's lemma (*cf.* [1], [2], [5], [7], [8], [9], [10], [11], [13]). It was stated and proved long before Zorn's lemma, is somewhat simpler in form, and often just as convenient to use. Like Zorn's lemma, it owes its great usefulness to the fact that it allows one to avoid the theory of ordinal numbers and well-ordered sets in giving proofs in abstract mathematics. For this reason it would be desirable to have a proof of the theorem which is independent of the notion of a well-ordered set, and dependent only on the axiom of choice. However, all the proofs in the literature, either of the maximal chain theorem or of Zorn's lemma, seem to involve the notion of a well-ordering (*cf.* [1], [2], [5], [6], [7], [8], [9], [11]). The following proof does not involve the notion of a well-ordered set. It was suggested by Zermelo's second proof of the well-ordering theorem in [12].

Theorem. *Every chain of a partially ordered set is contained in a maximal chain.*

Proof. A chain is a simply ordered set; that is, of any two distinct elements of a chain, one necessarily precedes the other. Suppose the theorem false. Then in some partially ordered set P there is a chain A not contained in any maximal chain. Then corresponding to each chain C which includes A as a subset select, by means of the axiom of choice, a larger chain C', called the successor of C, containing only a single element of P not in C. This is possible since by assumption no chain which includes A is maximal.

We shall call a collection K of chains of P all of which include A *complete* if A is in K, the successor of each member of K is in K, and K contains the union of each chain of its chains. Clearly the collection of *all* chains which include A is complete, and the intersection of any set of complete collections is complete. Let J be the intersection of *all* complete collections of chains of P. Then J is the smallest complete collection. We wish to prove that J is a chain, which will involve a contradiction.

* Received July 11, 1951.

We shall call a chain C which is a member of the collection J *normal* if for every chain X of J either $X \subset C$ or $C \subset X$. It will be proved that every member of J is normal. If C is any normal member of J, define $K(C)$ to consist of all members X of J such that either $X \subset C$ or $C' \subset X$. Now the collection $K(C)$ is complete, since in the first place A is in $K(C)$ since $A \subset C$. Secondly, if X is in $K(C)$ so is its successor X'. For by the definition of $K(C)$, either $C' \subset X$ or $X \subset C$. If $C' \subset X$, then $C' \subset X'$. On the other hand, if $X \subset C$, then either $X' \subset C$ or $C \subset X'$, since C is normal. If $X' \subset C$, then X' is in $K(C)$. But if $C \subset X'$, then $X \subset C \subset X'$, and since X' has only a single element of P not in X, it follows that $C = X$ or $C = X'$, whence $C' \subset X'$ or $X' \subset C$. In either case X' is in $K(C)$, which therefore contains the successor of each of its members. Likewise $K(C)$ clearly contains the union of each chain of its chains, since the defining property $X \subset C$ or $C' \subset X$ of the collection $K(C)$ goes over to unions. It follows that $K(C)$ is complete, and since it is a subset of J, the smallest complete collection, $K(C)$ must be identical with J.

But by the defining property of the collection $K(C)$, it follows that the successor C' of every normal member C of J is also normal. Since the union of a chain of normal chains is clearly normal, it is seen that the collection of all normal members of J is complete, and hence is identical with the collection J itself, since J is the smallest complete collection. Hence J is a chain of chains, since all of its members are normal. Since J is complete and also a chain, it must contain the union U of all of its members, and likewise it must contain the successor U' of U, which is a proper superset of U. But this is impossible. This contradiction proves the theorem.

Conclusion. Zorn's lemma, which states that every collection of sets which contains the union of each chain of its sets has a maximal element, is an immediate consequence of the maximal chain theorem (*cf.* [1], [8], [9], [13]). Another formulation of Zorn's lemma states that if every chain of a partially ordered set P has an upper bound in P, then P contains a maximal element (*cf.* [1], [8], [9]).

The well-ordering theorem and the axiom of choice may also be derived as consequences of the maximal chain theorem by defining properly a partial ordering between the well-ordered subsets of a given set, or between the selection functions defined on subsets of a given set.

THE PENNSYLVANIA STATE COLLEGE.

REFERENCES.

[1] Garrett Birkhoff, *Lattice Theory*, revised edition, Colloquium Publications 25, New York, 1948, pp. 42-44.

[2] N. Bourbaki, *Éléments des mathématiques, Théorie des ensembles I_1*, Paris, 1939, pp. 36-37.

[3] F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1914, pp. 140-141.

[4] G. Hessenberg, "Kettentheorie und Wohlordnung," *Journal für die reine und angewandte Mathematik*, vol. 135 (1908), pp. 81-133.

[5] H. Kneser, "Eine direkte Ableitung des Zornschen Lemmas aus dem Auswahlaxiom," *Mathematische Zeitschrift*, vol. 53 (1950), pp. 110-113.

[6] C. Kuratowski, "Une méthode d'élimination des nombres transfinis des raisonnements mathématiques," *Fundamenta Mathematica*, vol. 3 (1922), pp. 76-108.

[7] R. L. Moore, *Foundations of Point Set Theory*, Colloquium Publications 13, New York, 1932, pp. 84-85.

[8] Szele, T., "On Zorn's lemma," *Publicationes Mathematicae Debrecen*, vol. 1 (1950), pp. 254-257.

[9] J. W. Tukey, *Convergence and Uniformity in Topology*, Princeton, 1940, pp. 7-8.

[10] A. D. Wallace, "A substitute for the axiom of choice," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 278.

[11] E. Witt, "On Zorn's theorem," *Revista Matematica Hispano-Americana*, vol. 10, pp. 82-85 (1950).

[12] E. Zermelo, "Neuer Beweis für die Möglichkeit einer Wohlordnung," *Mathematische Annalen*, vol. 65 (1908), pp. 107-128.

[13] Max Zorn, "A remark on method in transfinite algebra," *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 667-670.

NOTES ON LEFT DIVISION SYSTEMS WITH LEFT UNIT.*

By M. F. Smiley.

1. R. Baer introduced in [3] the notion of *left division system with left unit* [2] by showing that these systems arise in a natural way from a simple method of multiplying the cosets of a subgroup of a group. It is our contention that many of the properties of loops [3] are valid also for these much more general systems. This is the first of a series of notes in which we shall support this contention. Herein we are concerned with basic structure theory culminating in the lemma of Zassenhaus. We owe the brevity [4] of the proofs in part to suggestions of a referee.

Let G be a system. We shall use $H \subseteq_s G$ as an abbreviation for "H is a subsystem of G." We observe that (Z1) $[e] \subseteq_s G$ *if* e *is the left unit of* G, and that (Z2) *the intersection of a family of subsystems of* G *is a subsystem of* G. We shall then denote the subsystem of G which is generated by H, $K \subseteq_s G$ by $\langle H, K\rangle$. If K is the kernel of a homomorphism η of G onto a system G', we shall write $K \subseteq_n G$. If $K \subseteq_n G$, we have (Z3) *if* $H \subseteq_s G$, *then* $H\eta \subseteq_s G'$, (Z4) *if* $H' \subseteq_s G'$, *then* $H'\eta^{-1} \subseteq_s G$, and (Z5) *if* $H \subseteq_n G$, *then* $H\eta \subseteq_n G' = G\eta$. In order to establish (Z5) we first prove that a subsystem K of G satisfies $K \subseteq_n G$ if and only if $(Kx)(Ky) = K(xy) = (Kx)y$ for every $x, y \varepsilon G$. (*Cf.* Baer [4], p. 455, Lemma 1.) When $K \subseteq_n G$, the mapping $x\eta \rightarrow Kx$ is an isomorphism of G', and $G/K = [Kx; x \varepsilon G]$.

Let us now mention a few immediate consequences of our definitions. If K is the kernel of a homomorphism η of G onto a system G', and $x\eta = y\eta$ for $x, y \varepsilon G$, then $x = ky$ for some $k \varepsilon K$. From (Z3) and (Z4) we see that $\langle H, K\rangle\eta = \langle H\eta, K\eta\rangle$ for $H, K \subseteq_s G$. We note that $K \subseteq_n G$ and $K \subseteq H \subseteq_s G$

* Received July 5, 1951; revised September 24, 1951.

[2] In the remainder of this note we shall use the word *system* in place of this longer phrase. Systems with unit were called *left loops* by Kiokemeister and Whitehead [13]. Their *admissible* left loops need not be normal in our sense, and there is no overlapping of our results for systems with theirs for left-loops.

[3] See [1, 2, 4-7, 10, 11].

[4] The use of the associative law will not shorten our proofs. It is, of course, well-known that an associative system is a group.

imply that $K \subseteq_n H$. The *modular law* $(KL \cap H) = (K \cap H)L$ for subsystems H, K, L of G with $L \subseteq H$ may be proved as in Baer [4].

The principal result of this note is the following lemma.

LEMMA. *If G is a system, $L \subseteq_n H \subseteq_s G$, and $K \subseteq_n G$, then* (1) $H \cap K \subseteq_n H$, (2) $KL \subseteq_n KH$, (3) $(KL \cap H) \subseteq_n H$, (4) *the mapping* $(KL \cap H)h \rightarrow (KL)h$ *with* $h \varepsilon H$ *is an isomorphism of* $H/(KL \cap H)$ *and* KH/KL, (5) *if we have* $H \subseteq_n G$, *then* $KH = HK \subseteq_n G$, *and* (6) *if we have* $K \cap H \subseteq L$, *then* $Lh \rightarrow (KL)h$ *with* $h \varepsilon H$ *is an isomorphism of* H/L *and* KH/KL.

Proof. Let η be a homomorphism of G onto a system G' with kernel K. By (Z3), $H\eta$ is a subsystem of G'. Thus η induces a homomorphism of H onto the system $H\eta$ and the kernel of this induced homomorphism is $H \cap K$. This proves (1). Now set $W = \langle H, K \rangle$. Then we have $W\eta = \langle H\eta, K\eta \rangle = H\eta$, and $W \subseteq KH$, $W = KH \subseteq_s G$. If also $H \subseteq_n G$, then $W = KH = HK$ (cf. Baer [4], p. 452). By (Z5), $L\eta \subseteq_n H\eta = (KH)\eta$. Let ϕ be a homomorphism of $H\eta$ onto a system H'' with kernel $L\eta$. Then $\eta\phi$ is a homomorphism of KH onto H'' with kernel KL. This proves (2). Again $\eta\phi$ induces a homomorphism of H onto H'' whose kernel is $KL \cap H$. This proves (3), which is also an immediate consequence of (1) and (2). If $h \varepsilon H$ and $k \varepsilon K$, we have $(kh)\eta\phi = h\eta\phi$, and it follows that $(KL)(kh) = (KL)h$, since $(hk)\eta\phi \rightarrow (KL)(kh)$ is an isomorphism of H'' and KH/KL. We then obtain (4) by noting that $h\eta\phi \rightarrow (KL \cap H)h$ is an isomorphism of H'' and $H/(KL \cap H)$. The statement (5) follows from (2) and our previous observation that $H \subseteq_n G$ implies $HK = KH$. Finally, (6) follows from (4) and the modular law.

COROLLARY (LEMMA OF ZASSENHAUS). *Let G be a system,*

$$A, B, A_1, B_1 \subseteq_s G, A_1 \subseteq_n A, B_1 \subseteq_n B.$$

Then

$$A_1(A \cap B_1) \subseteq_n A_1(A \cap B), \qquad B_1(B \cap A_1) \subseteq_n B_1(B \cap A),$$

and the identity mapping of G induces an isomorphism of the corresponding quotient systems.

Proof. Since $A \cap B \subseteq_s A$ and $A_1 \subseteq_n A$, (1) gives $A_1 \cap B \subseteq_n A \cap B$. Likewise, $B_1 \cap A \subseteq_n A \cap B$. Using (5), we see that

$$(A_1 \cap B)(A \cap B_1) = (A \cap B_1)(A_1 \cap B) = L \subseteq_n H = A \cap B.$$

We set $K = A_1$ and $G = A$, noting that $K \cap H = A_1 \cap B \subseteq L$, apply (6), and interchange A and B in the result.

Remarks. 1. If G is associative, then G is a group, and our discussion includes this case. On the other hand, if G is a loop, then $K \subseteq_n G$ does not imply that K is a normal subloop [5] in the sense of [1]. Thus our present exposition does not apply to loops.

2. However, it is possible to formulate a list of axioms which hold for systems and for loops and which justify our Lemma. We are indebted to R. Baer who suggested that such a list must exist. We consider a set $\mathfrak{J}$ of sets G with binary compositions which have a left unit e, for which $xa = b$ has a solution $x \varepsilon G$ for every $a, b \varepsilon G$, and such that $xa = a$ implies that $x = e$. If H is a subset of G which is an element of $\mathfrak{J}$ relative to the composition of G, we write $H \subseteq_t G$. Let $\mathfrak{Z}$ be a subset of $\mathfrak{J}$ such that the requirements (Z1)-(Z5) of our second paragraph hold. We agree, of course, to replace "system" by "element of $\mathfrak{Z}$" and "H is a subsystem of G" by "$H \subseteq_t G$ and H is an element of $\mathfrak{Z}$." We add the requirement: (Z6) If μ is a homomorphism of $G \varepsilon \mathfrak{Z}$ onto $G' \varepsilon \mathfrak{Z}$ with kernel K, then $K(xy) \subseteq (Kx)(Ky)$ for every $x, y \varepsilon G$. The interested reader will be able easily to adapt our proofs to these axioms if he adds the hypothesis $K \subseteq_n G$ to the statement of the modular law.

3. It is interesting to observe that a recent theorem of R. C. Buck [12] on homomorphisms is valid for multiplicative elements of the set $\mathfrak{J}$ of Remark 2. But trivial examples show that both our lemma and Buck's theorem fail in general.

4. Similarity (isotropy) [3, 1, 2, 11] will not reduce the study of systems to the study of loops. For if $G = [1, a, b]$, with 1 a left unit, $a1 = b$, $b1 = a$, $aa = ab = 1$, $ba = b$, $bb = a$, then G is a system which is not isotropic to a loop. The referee has remarked that every system is isotropic to a left loop; but, of course, not every left loop is a loop.

THE STATE UNIVERSITY OF IOWA.

[5] For a simple modification of the example of Bates and Kiokemeister [9] shows that a loop may have a system which is not a loop as a homomorphic image.

REFERENCES.

[1] A. A. Albert, "Quasigroups I," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 507-519.

[2] ———, "Quasigroups II," *ibid.*, vol. 55 (1944), pp. 401-419.

[3] R. Baer, "Nets and groups I," *ibid.*, vol. 46 (1939), pp. 110-141.

[4] ———, "The homomorphism theorems for loops," *American Journal of Mathematics*, vol. 67 (1945), pp. 450-460.

[5] ———, "Splitting endomorphisms," *Transactions of the American Mathematical Society*, vol. 61 (1947), pp. 508-516.

[6] ———, "Endomorphism rings of operator loops," *ibid.*, vol. 61 (1947), pp. 517-529.

[7] ———, "Direct decompositions," *ibid.*, vol. 62 (1947), pp. 62-97.

[8] ———, "Direct decompositions into infinitely many summands," *ibid.*, vol. 64 (1948), pp. 519-551.

[9] G. E. Bates and F. Kiokemeister, "A note on homomorphic mappings of quasigroups into multiplicative systems," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 1180-1185.

[10] B. Brown and N. H. McCoy, "Some theorems on groups with application to ring theory," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 302-311.

[11] R. H. Bruck, "Contributions to the theory of loops," *ibid.*, vol. 60 (1946), pp. 245-354.

[12] R. C. Buck, "A factoring theorem for homomorphisms," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 135-137.

[13] F. Kiokemeister and G. W. Whitehead, "A coset theory for left loops," (Abstract), *Bulletin of the American Mathematical Society*, vol. 51 (1951), pp. 60-61.

A CHARACTERIZATION OF FINITE DIMENSIONAL CONVEX SETS.*

By E. G. Straus and F. A. Valentine.

Let S be a *closed connected* set contained in a finite dimensional subspace of a linear space, and let R_n be the linear subspace of minimal dimension which contains S. A *maximal* convex subset of S is, by definition, one which is not contained in a larger convex subset of S. It is our purpose to establish the following result.

Theorem 1. *The set S defined above, is convex if and only if each point $x \,\varepsilon\, S$ is contained in a unique maximal convex subset $K(x)$ of S of dimension greater than or equal to $n-1$.* (Note: Observe that no restrictions are placed on the maximal convex subsets of S of dimension less than $n-1$).

Definition 1. The property placed on S in Theorem 1 is called property A. The symbol $|K(x)|$ denotes the maximal $(n-1)$-dimensional volume obtained from the class of all $(n-1)$-dimensional plane projections of $K(x)$. ($|K(x)|$ may be finite or infinite).

Lemma 1. *If there exists a non-convex closed connected set S with property A, then there exists a non-convex continuum*[1] *S_0 with property A, such that*

$$\operatorname*{g.l.b.}_{x \,\varepsilon\, S_0} |K(x)| \equiv d > 0.$$

Proof. Choose a point $x_1 \,\varepsilon\, S$, and let C_m be the solid sphere in R_n of radius m with center at x_1. Also let $F_m \equiv S \cdot C_m$, and denote the component of F_m which contains x_1 by T_m. Since $\sum_{m=1}^{\infty} T_m = S$, and since S is non-convex, there exists a fixed value m such that T_m is not convex. Consider now the non-convex set T_{m+1}. In a well-established manner (*cf.* [2], Ch. 7), we may determine a topological space each element of which corresponds to one and only one of the maximal convex subsets of T_{m+1} of dimension $\geqq n-1$. By this process of regarding maximal convex subsets of T_{m+1} as points in a new

* Received August 20, 1951.

[1] A continuum is a compact connected set.

space, it is clear that T_{m+1} is mapped into a *set* called T^*_{m+1}, and it is also easily verified that T_m is mapped into a non-trivial *subcontinuum* T^*_m of T^*_{m+1}. (It should be noted that although T^*_m will be closed, bounded and connected, T^*_{m+1} may not be closed). To each point $x^* \varepsilon T^*_m$, we can assign the positive valued function $f(x^*) \equiv | x^* | \equiv | K(x) |$, where $K(x)$ is the maximal convex subset of T_{m+1} corresponding to x^*. This function $f(x^*)$ is upper semicontinuous on T^*_m. Hence the set $S^*_k \equiv \{x^* \mid f(x^*) \geqq 1/k, x^* \varepsilon T^*_m\}$ is closed. If S^*_k were nowhere dense in T^*_m for each k, then $T^*_m = \sum_{k=1}^{\infty} S^*_k$ would be of the first category in itself, which contradicts the fact that it is a non-trivial continuum. Hence, there exists a value k such that S^*_k contains a non-trivial subcontinuum S^*_0. The pre-image S_0 of S^*_0 must then be a non-convex continuum in T_{m+1}, since S_0 is a closed connected subset of the compact set T_{m+1}. Moreover, S_0 satisfies property A, since each point $x \varepsilon S_0$ lies in a unique maximal convex subset $K(x)$ of T_{m+1} of dimension $\geqq n-1$ which lies in $S_0(K(x)$ must intersect $T_m)$. Also for each $x \varepsilon S_0$, we have $| K(x) | \geqq 1/k$, so that $d = \text{g.l.b.}_{x \varepsilon S_0} | K(x) | \geqq 1/k > 0$.

LEMMA 2. *If S_0 and d are the quantities defined in Lemma 1, then*

$$\underset{x \varepsilon S_0,\, \dim K(x)=n-1}{\text{g.l.b.}} | K(x) | = d.$$

Proof. The sets $K(x)$ with dimension $n-1$ are everywhere dense in S^*_0 since property A implies there exists at most a denumerable number of $K(x)$ with dimension n. This combined with the upper semicontinuity of $| K(x) |$ proves the lemma.

LEMMA 3. *Each plane P through the centroid of a bounded convex set K of dimension m and volume V divides K into two convex sets whose volumes V_1 and V_2 satisfy the inequalities*

$$V_i \geqq c_m V \qquad (i = 1, 2),$$

where c_m is a constant depending solely on m and not on K or P.

This lemma was proved for $m = 2$ by Neumann in [1], and for arbitrary m by Green and Gustin [unpublished].

LEMMA 4. *Consider the set S_0 in Lemma 1, and let $K(x)$ be a maximal convex subset of dimension $n-1$ such that $| K(x) | < (1 + c_{n-1})d$, where c_{n-1} is defined in Lemma 3. Let T_r be the right circular solid cylinder of radius r, whose axis passes through the centroid of $K(x)$ and is perpendicular to the plane of $K(x)$.*

Then there exists a neighborhood U of $K(x)$ and a value of r such that for each point $y \in U \cdot S_0$, the set $K(y)$ intersects all the elements of T_r.

Proof. If the lemma were false, there would exist a sequence of convex sets $K(y_i)$ whose distance from $K(x)$ would approach zero, and such that the following would hold. The sequence $K(y_i)$ would contain a subsequence which would converge to a convex subset C of $K(x)$ which does not contain the centroid of $K(x)$ in its interior. Due to the upper semicontinuity of $|K(x)|$, we have $|c| \geqq d$. Hence, the above with Lemma 3 implies that $|K(x)| \geqq |c| + c_{n-1}d \geqq (1 + c_{n-1})d$, which contradicts our hypothesis.

Sufficiency proof for Theorem 1. Assume that S is not convex, so that Lemmas 1 to 4 hold. We use the notation in Lemma 4. According to a known theorem in topology ([2], p. 16), there exists a non-trivial component C of $S_0 \cdot U$ containing $K(x)$. Since S_0 is non-convex, C is non-convex. Choose a point $y_1 \in C$. Since $K(y_1) \cdot T_r$ is a convex set dividing T_r into two parts, let L_1 be *any* line segment parallel to the axis of T_r, and having its endpoints in $K(y_1) \cdot T_r$ and $K(x)$ respectively. Let L be the line containing L_1. For any point $y \in C$, Lemma 4 implies $K(y) \cdot L \neq 0$. Hence, since C is a connected closed bounded set, property A implies that $C \cdot L$ is a closed line segment. Hence $L_1 \subset C$. But this implies that the portion of T_r between $K(x)$ and $K(y_1)$ is in C. However, this contradicts the fact that $K(x)$ is of dimension $n-1$. This completes the sufficiency. The necessity of property A is obvious.

An interesting corollary to Theorem 1 for sets in R_2, the plane, is the following.

COROLLARY 1. *If S is a closed connected set in R_2, each point of which belongs to a unique maximal linear element (line segment, half-line or line) of S, then S is a linear element.*

Concluding remarks. One may ask in what respects our theorem is the best possible. It is the best possible in at least the following respects.

If we remove the assumption $\dim K(x) \geqq n-1$, then the theorem no longer holds, as shown by the circular cylinder $x_1^2 + x_2^2 = 1$ in E_n.

It might be conjectured that our affine theorem has a purely topological origin. Thus one might think that if we replace the phrase "unique maximal convex subset of dimension $\geqq n-1$" by "unique maximal closed (in the point set sense) surface of dimension $n-1$" then the conclusion would be "S is an $n-1$ dimensional closed surface." However, P. Erdös and A. H.

Stone have communicated to us the following counterexample for the case $n = 2$. Consider a Cantor set in the interval $(0, 1)$ of the X-axis. On each of the points of this set we erect a line segment of unit length perpendicular to the X-axis on which $y \geqq 0$. Consider the intervals in the complement of the Cantor set. For each interval I, draw the diagonal segment joining the upper end point of the vertical segment at the left endpoint of I to the right endpoint of I. The resulting set S of vertical segments plus diagonal segments is connected, and every point in S belongs to a unique maximal closed arc (either a vertical segment or a polygonal line consisting of two vertical segments plus a diagonal segment). However, S is *not* an arc.

The following question is still unsettled. Suppose $S \subset R_n$ is a closed connected set such that each of its points is contained in a unique maximal closed connected $(n-1)$-dimensional subset of S which is contained in an $(n-1)$-dimensional hyperplane of R_n. Is then S an $(n-1)$-dimensional set which is contained in an $(n-1)$-dimensional hyperplane? If $n = 2$, the question is answered in the affirmative by Corollary 1. For $n > 2$, the question is still undecided.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES.

REFERENCES.

[1] B. H. Neumann, "On an invariant of plane regions and mass distributions," *Journal of the London Mathematical Society*, vol. 20 (1945), pp. 226-237.

[2] G. T. Whyburn, *Analytic Topology*, American Mathematical Society Colloquium Publications (1942).

ON ADDITIVE IDEAL THEORY IN GENERAL RINGS.[1]

By Charles W. Curtis.

Introduction. It is the purpose of this paper to present some contributions to the structure theory of non-commutative ideal lattices, as developed by Krull [10][2] and Dilworth [3]. The results of Part 1, except for (5) of Lemma 1.1 and the first part of Theorem 1.4, hold in an arbitrary non-commutative residuated lattice in the sense of Dilworth [3]. The results are stated only for ideals[3] in rings, however, since our interest is the application of these results to the structure theory of rings. The main result of the paper is the determination in **1.4** of the maximal elements in the inclusion ordered set of right associated primes of an ideal. In **1.1**, **1.2** **1.3** a decomposition theory of ideals is worked out, similar to the work of Fuchs [5] for commutative rings. In **1.5** Krull's theory [10] of right associated prime ideals is linked to the theory of primary ideals. The notion of isolated component ideal leads to a new approach to the uniqueness theory of primary ideals. In **1.6** it is proved that if the ring R satisfies the ascending chain condition (A. C. C.) for right ideals, and if every ideal in R is a finite intersection of primary ideals, then the intersection of the powers of the Jacobson radical is the zero ideal. The methods presented here do not lead to a proof of this theorem for arbitrary rings with A. C. C. for right ideals because of an example due to E. Noether, published in [10], of a ring with A. C. C. for right ideals having the property that not every ideal is an intersection of primary ideals.

I wish to thank Professor N. Jacobson for the encouragement and many helpful suggestions he has given me during the preparation of this paper.

Part 1. The General Theory.

1.0. Notations. In this paper R will denote a non-commutative ring with a unit element 1. Proper ideals in R will be denoted by $A, B, C, \cdots$, and elements of R by $x, y, a, b, \cdots$. We shall use the symbol 0 both for

[1] Received June 29, 1951.

[2] Numbers in brackets refer to the list of references at the end of the paper.

[3] "ideal" always means "two-sided ideal."

the zero ideal and for the zero element of R. $\{\Sigma C \mid C$ has the property $P\}$ means the join of those ideals C having the property P, while $\{x \mid x$ has the property $P\}$ means the set of all elements x of R having the property P.

1.1. Primal ideals.

Definition 1.1. The *quotient* AB^{-1} of the ideals A and B is defined by $AB^{-1} = \{\Sigma C \mid CB \subseteq A\}$. Similarly $B^{-1}A = \{\Sigma C \mid BC \subseteq A\}$.

We observe that $AB^{-1} = \{x \mid xB \subseteq A\}$; $B^{-1}A = \{x \mid Bx \subseteq A\}$. The following properties of the quotients are well known ([10], [3]).

LEMMA 1.1.

(1) $(AB^{-1})B \subseteq A$, $AB^{-1} = R$ *if* $B \subseteq A$, *and if* $B_1 \subseteq B_2$ *then* $AB_1^{-1} \supseteq AB_2^{-1}$.

(2) $(AC^{-1})B^{-1} = A(BC)^{-1}$.

(3) $(A_1 \cap \cdots \cap A_n)B^{-1} = A_1B^{-1} \cap \cdots \cap A_nB^{-1}$,

(4) $A(B_1 + \cdots + B_n)^{-1} = AB_1^{-1} \cap \cdots \cap AB_n^{-1}$, *and more generally, if* $\{B_\mu\}$ *is an arbitrary collection of ideals,*

(5) $A(\Sigma_\mu B_\mu)^{-1} = \bigcap_\mu AB_\mu^{-1}$.

Similar rules hold for the quotients $B^{-1}A$.

Definition 1.2. If $AB^{-1} = A$ $(B^{-1}A = A)$ then B *is right* (left) *prime* to A.

Definition 1.3. A is *right primal*[4] if the join P of the ideals not right prime to A is again not prime to A. If A is primal, then P is called the *adjoint ideal* of A.

Let R be a commutative ring. If G is primal in the sense of Definition 1.3 then G is primal in the sense of Fuchs [5]. The converse is false.[5]

---

[4] Although there is also a theory of left primal ideals, we shall consider only right primal ideals. Henceforth, "primal" means "right primal," and "prime to A" means "right prime to A."

[5] This point is settled by the following example, due to the referee. Let R be the set of all finite expressions $a_1t^{n_1} + \cdots + a_rt^{n_r}$, with coefficients in a field, and exponents non-negative rational numbers. With respect to the obvious definitions of addition and multiplication, R is a commutative ring in which the principal ideal generated by t is primal in the sense of Fuchs, but not primal in the sense of Definition 1.3.

Definition 1.4. P is a *prime ideal*[6] in an arbitrary ring R, if $AB \subseteq P$ implies $A \subseteq P$ or $B \subseteq P$.

From Lemma 1.1, (2) we have

LEMMA 1.2. *If A is primal, then the adjoint ideal P of A is a prime ideal.*

Definition 1.5. A is *strongly irreducible* if A cannot be expressed as an intersection, finite or infinite, of proper divisors of A. A is *irreducible* if A cannot be expressed as an intersection of a finite number of proper divisors of A.

From Lemma 1.1, (5) and (4), we obtain at once

LEMMA 1.3. *Every strongly irreducible ideal is right primal. If R satisfies the A.C.C. for ideals, then every irreducible ideal is right primal.*

THEOREM 1.4. *Every ideal in R is the intersection of its primal divisors. If R satisfies the A.C.C. for ideals, then every ideal in R is an intersection of a finite number of primal ideals.*[7]

Proof. Let A be an ideal in R. Since there are no ideals not prime to the ring R itself, we shall agree that R is a primal ideal. In order to prove the first part of the theorem it is sufficient to prove that if $c \not\in A$, then there exists a primal divisor G of A not containing c. From Zorn's lemma, however, it follows that there exists a divisor G of A having the property that every proper divisor of G contains c. G is therefore strongly irreducible, and by Lemma 1.3, primal. The second part of the theorem is an immediate consequence of the A.C.C. and Lemma 1.3.

1.2. Uniqueness of primal decompositions.

Definition 1.6. (E. Noether [17]) The intersection $A = G_1 \cap \cdots \cap G_n$ is *irredundant* if no G_i divides its *complement*

$$G_1 \cap \cdots G_{i-1} \cap G_{i+1} \cap \cdots \cap G_n.$$

---

[6] This definition is due to Krull [10].

[7] The first right principal components defined by Krull in [10] are primal ideals, and therefore a theorem of [10], which states that every ideal is the intersection of its first right principal components, is an instance of Theorem 1.4. I am indebted to the referee for pointing out to me that since an ideal A is strongly irreducible if and only if the ring R/A is subdirectly irreducible in the sense of Birkhoff ("Subdirect unions in universal algebra," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 764-768), Theorem 1.4 is a consequence of Birkhoff's result, which states that an arbitrary ring is a subdirect sum of subdirectly irreducible rings.

The intersection is *reduced* if no G_i can be replaced by a proper divisor. Intersections which are both irredundant and reduced are called *normal.*

Throughout the remainder of this section we shall assume that R satisfies the A. C. C. for ideals. The idea of the next result is due to Fuchs [5].

THEOREM 1.5. *Let $A = G_1 \cap \cdots \cap G_n$ be a reduced representation*[8] *of A by primal ideals G_i with adjoint prime ideals P_i. An ideal B is not prime to A if and only if B is contained in one of the P_i.*

Proof. By (4) we have

$$AB^{-1} = (G_1 \cap \cdots \cap G_n)B^{-1} = G_1B^{-1} \cap \cdots \cap G_nB^{-1}.$$

Since the intersection is reduced, $AB^{-1} \neq A$ if and only if $G_iB^{-1} \neq G_i$ for some i. But $G_iB^{-1} \neq G_i$ if and only if $B \subseteq P_i$, and the proof is complete.

THEOREM 1.6. *The reduced intersection of a finite number of primal ideals $A = G_1 \cap \cdots \cap G_n$ with adjoint ideals $P_1, \cdots, P_n$ is primal if and only if one prime P_j divides all the others.*

Proof. First let some P_i, say P_1 divide all the other P_j. By Theorem 1.5, if $AB^{-1} \neq A$, then $B \subseteq P_i \subseteq P_1$ for some i; hence if S denotes the join of the ideals not prime to A, $S \subseteq P_1$. But again by Theorem 1.5 we conclude that $AS^{-1} \neq A$, and hence A is primal.

Conversely, let S be the adjoint ideal of A. Since $AS^{-1} \neq A$, $S \subseteq P_i$ for some i. But $\sum_{j=1}^{n} P_j \subseteq S \subseteq P_i$, and the theorem is proved.

THEOREM 1.7. *Every ideal in R is a normal intersection of primal ideals, such that no adjoint prime divides another.*

Proof. By the A. C. C. A is an intersection of irreducible ideals, and we can assume that the intersection is irredundant. By Lemma II of E. Noether's paper [17] which holds, together with the other results we shall require from that paper, in the non-commutative case, the intersection is necessarily reduced. By Lemma 1.3, the ideals appearing in the intersection are primal ideals. Let their adjoint primes be $P_1, \cdots, P_n$, and suppose the indices chosen so that $P_1, \cdots, P_k$ are the maximal elements in the set $\{P_i\}$, ordered by inclusion. If we replace the intersection G_1 of those primal ideals whose adjoint primes are P_1 or a multiple of P_1 by G_1 itself, then G_1 is, by Theorem 1.6, a primal

[8] By a representation of A, we mean an expression of A as an intersection of some of its divisors.

ideal, and by Lemma IV of [17] the resulting intersection is still reduced. Next we replace the intersection G_2 of those ideals, not already incorporated into G_1, whose adjoint ideals are P_2 or a multiple of P_2 by G_2, and again by Lemma IV of [17] the intersection is reduced. By repeating this process k times, we obtain at last a normal representation of A by primal ideals having the desired properties.

THEOREM 1.8. *If A has two normal representations by primal ideals, such that no adjoint prime in either representation divides another in the same representation, then the adjoint primes in the two representations and the number of components are the same.*

Proof. Let $A = G_1 \cap \cdots \cap G_n = G_1^* \cap \cdots \cap G_m^*$ be two normal primal representations of A satisfying the hypotheses of the theorem. Let the adjoint primes be $P_1, \cdots, P_n$ and $P_1^*, \cdots, P_m^*$. We shall prove that P_1 is contained in some P_j^*. Since both representations are reduced, we can apply Theorem 1.5 once to conclude that P_1 is not prime to A, and again to conclude that P_1 is contained in some P_j^*. By symmetry we can show that $P_j^* \subseteq P_k$ for some k. Thus $P_1 \subseteq P_j^* \subseteq P_k$, which contradicts our hypothesis unless $P_1 = P_j^* = P_k$. The rest of the proof is now clear.

1.3. Maximal prime ideals. McCoy defined a set S in a ring R to be an m-system if $a, b \varepsilon S$ imply the existence of an element $x \varepsilon R$ such that $axb \varepsilon S$. The empty set is considered to be an m-system. He then defined the *radical* $M(A)$ of an arbitrary ideal A in R to be the set of elements r such that every m-system containing r contains an element of A. McCoy proved ([15], Theorem 2) that the radical $M(A)$ of an ideal A is the intersection of the minimal prime divisors of A, thus achieving a successful generalization of the work of Krull [11]. Levitzki has sharpened this result by proving in [14] that $M(A)/A$ is actually the *lower radical* in the sense of Baer [2] of R/A.

For commutative rings, Fuchs has characterized in [6] the structure of the intersection of all maximal prime ideals belonging to an arbitrary ideal. For non-commutative rings R satisfying the A. C. C. for ideals, we shall prove a result analogous to Fuchs' theorem.

Definition 1.7. Let P be a divisor of A, where $A \neq R$. P is a *maximal prime ideal* belonging to A if (i) $AP^{-1} \neq A$, and (ii) if Q is a proper divisor of P, then $AQ^{-1} = A$. P is a *minimal prime ideal* of A if P is a prime ideal, and if there exists no prime ideal Q such that $A \subseteq Q \subseteq P$.

It should be observed that any ideal satisfying (i) and (ii) is necessarily a prime ideal, and consequently the definition is meaningful in the form in which we have stated it.

In order to conclude that A has at least one maximal prime ideal, it is enough to assume the A. C. C. for ideals. *Throughout this section we shall assume the A. C. C. for ideals.*

Definition 1. 8. The join S of all ideals C such that $A(C+B)^{-1} \neq A$ whenever $AB^{-1} \neq A$ is called the *adjoint ideal* of A (compare Fuchs [6]).

Since $A(A+B)^{-1} = AA^{-1} \cap AB^{-1} = R \cap AB^{-1} = AB^{-1}$, we see that $A \subseteq S$. Until now we have spoken of the adjoint ideal only in connection with primal ideals. We shall prove that if A is a primal ideal with adjoint prime S', then $S = S'$. In fact if $c \in S$, then (c)[9] is certainly not prime to A; hence $(c) \subseteq S'$, and we conclude that $S \subseteq S'$. Conversely, let $AB^{-1} \neq A$. Then $A(S'+B)^{-1} = A(S')^{-1} \neq A$; hence $S' \subseteq S$.

THEOREM 1. 9. *The adjoint ideal of A is the intersection of all maximal prime ideals of A.*

Proof. Let c be contained in the adjoint ideal S of A, and suppose that a maximal prime ideal P of A does not contain c. Then $P+(c)$ is a proper divisor of P, and since $AP^{-1} \neq A$, $A(P+(c))^{-1} \neq A$, contrary to our assumption that P is a maximal prime belonging to A. Conversely, let d be contained in every maximal prime of A, and let B be any ideal not prime to A. Then by the A. C. C. B is contained in a maximal prime, say P^*, and $B+(d) \subseteq P^*$; hence $B+(d)$ is not prime to A. Thus (d) and hence d is contained in the adjoint ideal of A.

By virtue of a remark of McCoy ([15], page 829) any prime ideal containing A contains a minimal prime of A. Since the McCoy radical is the intersection of all the minimal primes of A, it follows from this remark and Theorem 1. 9 that the adjoint ideal of A always contains the McCoy radical.

We have defined the set of maximal primes ideals belonging to A independently of a particular representation of A by primal ideals. The next theorem shows the connection between maximal prime ideals and primal representations.

THEOREM 1. 10. *Let $A = G_1 \cap \cdots \cap G_n$ be a normal primal decomposition of A, with adjoint primes $P_1, \cdots, P_n$. Then an ideal P is a maximal prime ideal belonging to A if and only if P is one of the P_i.*

---

[9] (c) is the principal ideal RcR generated by c.

Proof. We shall prove first that the P_i are maximal primes belonging to A. By Theorem 1.5, $AP_i^{-1} \neq A$ for each i, and again by Theorem 1.5, since the representation is reduced, each divisor of P_i is prime to A, $1 \leq i \leq n$. Consequently the P_i are maximal primes belonging to A. Conversely if P is any maximal prime belonging to A, then $AP^{-1} \neq A$, and by Theorem 1.8, $P \subseteq P_i$ for some i. From the maximality of P and the fact that $AP_i \neq A$, we conclude that $P = P_i$.

COROLLARY 1.11. *A is primal if and only if A has exactly one maximal prime ideal.*

1.4. Associated prime ideals. In the additive ideal theory of commutative rings it is desirable to give a definition of the prime ideals "associated" with a given ideal, independently of the notion of primary ideal. Krull gave such a definition for commutative rings in [9], and for non-commutative rings in [10]. We shall follow his methods in this section.

Definition 1.9. The ideal I is a (right) *isolated component ideal* (I. C. I.) of A if there exists an ideal B and an integer $q > 0$ such that $I = AB^{-q} = AB^{-q-1} = \cdots$.

We shall assume throughout this section that R satisfies the A. C. C. for ideals. If A is a given ideal, and B an arbitrary ideal, then we have in general $AB^{-1} \subseteq AB^{-2} \subseteq \cdots$, and by the A. C. C. there exists an integer q such that $AB^{-q} = AB^{-q-1} = \cdots$; the I. C. I. AB^{-q} is called the I. C. I. *generated* by B, and we shall denote it by $I(A, B)$, or more simply, when it is clear from the context that A is the basic ideal in the discussion, by $I(B)$.

Definition 1.10. A prime ideal P is a (right) *associated prime ideal* of A if (i) $I(P)$ is a proper divisor of A, and (ii) $I(P)^{-1}A \subseteq P$.

LEMMA 1.12. *Every maximal prime ideal P of A is a right associated prime ideal of A.*

Proof. Since $AP^{-1} \neq A$, $I(P)$, which contains AP^{-1}, is a proper divisor of A. It remains to prove that $I(P)^{-1}A \subseteq P$, that is, if C is an ideal such that $I(P)C \subseteq A$ then $C \subseteq P$. Let $D = AP^{-1}$; then D is a proper divisor of A, and $DP \subseteq A$. Since $D \subseteq I(P)$, we have also $DC \subseteq A$. Combining these results we obtain $D(P + C) \subseteq A$, and consequently $A(P + C)^{-1} \neq A$. Since P is a maximal prime of A, $P + C \subseteq P$, and we have $C \subseteq P$.

LEMMA 1.13. *Every right associated prime of A is contained in a maximal prime of A.*

Proof. This result follows immediately from the fact that if P is a right associated prime of A, then $AP^{-1} \neq A$.

From Lemmas 1.12 and 1.13 we obtain at once

THEOREM 1.14. *A prime ideal P is a maximal prime ideal of A if and only if P is a maximal element in the inclusion-ordered set of right associated prime ideals of A.*

1.5. Primary ideals. In this section we assume that R satisfies the A. C. C. for ideals.

Definition 1.11. An ideal Q is (right) *primary* if for arbitrary ideals A and B, $AB \subseteq Q$, $A \not\subseteq Q$ implies $B^r \subseteq Q$ for some positive integer r. As usual "primary" means "right primary" from now on. If Q is primary then the ideal

$$P = \{\Sigma B \mid B^r \subseteq Q \text{ for some positive integer } r\}$$

is called *the prime ideal belonging to Q*, or simply *the prime ideal* of Q.

It is easy to verify that since R satisfies the A. C. C., the ideal P defined above actually is a prime ideal.

If Q is primary, then Q is primal with adjoint prime P. The converse is false, as an example given in [5], page 2 indicates.

Let A be an ideal which is a finite intersection

$$A = Q_1 \cap \cdots \cap Q_k \tag{6}$$

of primary ideals. Dilworth has observed ([3], Theorem 6.1) that the intersection (6) can be refined by uniting those ideals Q_i having the same prime, obtaining an expression for A as an irredundant intersection $Q_1^* \cap \cdots \cap Q_m^*$ of primary ideals Q_i^*, where $Q_i^* \cap Q_j^*$ is not primary if $i \neq j$. Such an intersection we shall call a *shortest representation of A by primary ideals*, or more briefly, a S. R of A. Dilworth's result also states that for any S. R. of A, the primes and the number of primary ideals are uniquely determined. These results have also been announced by Murdoch [16].

Let $A = Q_1 \cap \cdots \cap Q_k$ be a S. R. of A by primary ideals Q_i with primes P_i. Consider a subset S of the set $\{P_i\}$ having the property that if $P_i \in S$, then $P_j \subseteq P_i$ implies $P_j \in S$. The intersection of the primary ideals belonging to the primes in S is called a *Noether isolated component* of A.

The methods of Krull [9] lead directly to a proof of the next result.

THEOREM 1.15. *If B is an ideal, then $AB^{-1} = AB^{-2}$ if and only if*

$A_1 = AB^{-1}$ is the ring R or a Noether isolated component of A appearing in every S. R. of A. Furthermore, if A_1 is any Noether isolated component of A, there exists an ideal B such that $A_1 = AB^{-1}$.

COROLLARY 1.16. *A_1 is a right isolated component ideal of A if and only if A_1 is a Noether isolated component of A.*

The following theorem is an application of Theorem 1.15 and the methods of Krull [9]. Since the previous theorem can be proved independently of Theorem 6.1 of [3], Part (b) of the next result (which is an immediate corollary of Part (a)) gives an alternative approach to Theorem 6.1 of [3].

THEOREM 1.17.

(a) *The set of right associated primes of A is identical with the set of prime ideals belonging to the primary ideals in every S. R. of A.*

(b) *The number of primary ideals in a S. R. of A and the primes belonging to them are uniquely determined.*

(c) *The Noether isolated components of A are uniquely determined by their associated primes.*

1.6. On the powers of an ideal. Let R be a ring satisfying the A. C. C. for right ideals, and having the property that every ideal in R is an intersection of a finite number of primary ideals. If A is an ideal in R, let $A^\omega = \bigcap_{i=1}^{\infty} A^i$. The first part of the proof of Satz 1 of [12] can be transferred to a ring satisfying the above hypotheses, proving

LEMMA 1.18. $A^\omega A = A^\omega$.

THEOREM 1.19. *If J is the Jacobson radical of R then $J^\omega = 0$.*

Proof. Lemma 1.18 and a result of Jacobson [8, Theorem 10].

Part 2. On Some Examples.

2.1. We shall say that R has a *Noetherian ideal theory* if i) R satisfies the A. C. C., and ii) every ideal in R is an intersection of a finite number of (right) primary ideals. The following example, mentioned in the introduction, shows that not every ring with A. C. C. has a Noetherian ideal theory. Let K be the field of rational numbers, and let R be the algebra over K with

basis elements e_1, e_2, n and the multiplication table $e_i^2 = e_i$, $i = 1, 2$; $n^2 = 0$; $e_1e_2 = e_2e_1 = 0$; $e_1n = ne_2 = n$; $e_2n = ne_1 = 0$. A straightforward determination of the ideals in R leads to a verification of the fact that the zero ideal of R is neither primary nor an intersection of primary ideals.

In view of this example and the fact that the results of 1.5 and 1.6 are valid for rings with a Noetherian ideal theory, it is important to consider examples of rings having a Noetherian ideal theory. It follows easily from the fundamental homomorphism theorem that if R has a Noetherian ideal theory, so has any homomorphic image of R. Another class of examples is furnished by finite matrix rings over rings with a Noetherian ideal theory.

From the results of Fitting [4] it follows directly that if R is a ring satisfying the A. C. C. for ideals, and having the property that every right or left ideal is a two-sided ideal, then R has a Noetherian ideal theory. This class of rings contains all Noetherian rings.[10]

Let R be a ring with A. C. C. for ideals, satisfying the conditions that if P is a prime ideal in R, different from the zero ideal, then P is maximal and if P_1 and P_2 are prime ideals in R, then $P_1P_2 = P_2P_1$. We shall apply the theory of primal ideals to prove that R has a Noetherian ideal theory. It is sufficient to prove that if A is an ideal in R (possibly the zero ideal), then A is a finite intersection of primary ideals. Let $A = Q_1 \cap \cdots \cap Q_n$ be a representation of A by primal ideals Q_i with adjoint prime ideals P_i. We shall prove that each Q_i is primary. In fact, consider Q_1. Let $\bar{P}_1$ be a minimal prime ideal of Q_1. Since the prime ideals commute, it follows from [10], Theorem 6 that $Q_1\bar{P}_1^{-1} \neq Q_1$, and hence $\bar{P}_1 \subseteq P_1$. Either $\bar{P}_1 = 0$ and hence $A = 0$, which shows that A is prime, or $\bar{P}_1 = P_1$. In the latter case Q_1 has the unique maximal and minimal prime P_1, and it follows that Q_1 is primary. Similarly the ideals $Q_2, \cdots, Q_n$ are primary. Examples of rings of this type are non-commutative principal ideal rings, and more generally, orders satisfying the axioms of Asano.[11]

We shall digress for a moment to consider other types of ideals related to primary ideals. We say that Q is *strongly right primary* if $ab \in Q$, $a \notin Q$ imply $b^r \in Q$ for some positive integer r.

In the commutative case, this definition coincides with Definition III given by E. Noether in [17], while our Definition 1.11 coincides with Definition IIIa of E. Noether. For commutative rings the two definitions

[10] A Noetherian ring is a commutative ring with unit element, satisfying the A. C. C. for ideals.

[11] Cf. [7], Chapters III and VI.

are equivalent if R satisfies the A. C. C., but this is no longer true for non-commutative rings.

Let R satisfy the A. C. C. for right ideals. If Q is strongly primary, then Q is primary, and we shall give an example to show that the converse is false. Let Q be strongly primary, and suppose $AB \subseteq Q$, $A \not\subseteq Q$, where A, B are ideals in R. By definition B/Q is a nil ideal in the ring R/Q, which also satisfies the A. C. C. for right ideals. By a result of Levitzki [13] every nil ideal in R/Q is nilpotent, and this proves that Q is primary. For the example, let D be a finite dimensional central simple algebra, and let $R = D_n$, the ring of n by n matrices with coefficients in D, for $n > 1$. R is a simple algebra, and its zero ideal is prime, and a fortiori primary. It is easy to find zero divisors in R.

Murdoch [16] has defined an ideal Q to be primary if $aRb \subseteq Q$, $a \not\in Q$ implies $b \in M(Q)$, where $M(Q)$ is the McCoy radical of Q. We shall prove that for rings satisfying the maximum condition for right ideals, Murdoch's definition is equivalent to Definition 1. 11. In fact, let Q be primary according to 1. 11, and let $aRb \subseteq Q$, where $a \not\in Q$. It follows that the ideal RbR is nilpotent modulo Q; hence $RbR \subseteq M(Q)$, and $b \in M(Q)$, since $M(Q)$ is a radical ideal. Conversely let Q be primary in the sense of Murdoch, and let $AB \subseteq Q$, $A \not\subseteq Q$ where A and B are ideals. It follows that B is a nil ideal modulo Q, and by Levitzki's result again, B is nilpotent modulo Q.

2. 2. Let S be a simple ring with unit element. Then it is well known that the center Φ of S is a field, and S is a central simple algebra over Φ. Let $S[X_1, \cdots, X_n]$, or more briefly $S[X]$, denote the algebra over Φ of polynomials in n indeterminates X_i, where we assume that the indeterminates are commutative with one another and with the elements of S. With the same assumptions concerning the X_i, let $S\{X_1, \cdots, X_n\}$, or more briefly $S\{X\}$, denote the algebra over Φ of formal power series in the variables X_i. Let R be either $S[X]$ or $S\{X\}$. In this section we shall prove that if either $R = S[X]$ and S is arbitrary, or if $R = S\{X\}$ and $(S:\Phi) < \infty$, then R has a Noetherian ideal theory.

Let M be the monomial basis for R, that is M consists of all monomials $X_1^{e_1}X_2^{e_2} \cdots X_n^{e_n}$ where the e_i are non-negative integers. Let $f \in R$. Then f can be written uniquely in the form $f = \Sigma a(m)m$, $a(m) \in S$, $m \in M$, where $a(m)$ is always a finitely valued function if $R = C[X]$, and $a(m)$ is an arbitrary function if $R = S\{X\}$. Let $B = (b, c, \cdots)$ be a basis for S over Φ. Then $a(m)$ can be written uniquely as $a(m) = \Sigma\lambda(m, b)b$, $\lambda(m, b) \in \Phi$, where the functions $\lambda(m, b)$ are b-finitely valued, that is the matrix $(\lambda(m, b))$ is row finite. We have

(1) $$f = \Sigma_m(\Sigma_b\lambda(m, b)b)m = \Sigma_{m,b}\lambda(m, b)bm$$

uniquely, and conversely if $(\mu(n, c))$ is any row finite matrix on $M \times B$ to Φ, $g = \Sigma\mu(n, c)cn$ is an element of R. Let $bc = \Sigma_d\xi(b, c, d)d$, and $mn = \Sigma_p\eta(m, n, p)p$ (for each b, c, $\xi(b, c, d)$ is d-finitely valued, and for each m, n, $\eta(m, n, p)$ is p — 1-valued.) Then

(2) $$\begin{aligned} fg &= (\Sigma\lambda(m, b)bm)(\Sigma\mu(n, c)cn) \\ &= \sum_{(m,n,b,c,d,p)} \lambda(m, b)\mu(n, c)\xi(b, c, d)\eta(m, n, p)dp, \end{aligned}$$

where the matrix

$$(\theta(p, d)) = (\sum_{(m,n,b,c)} \lambda(m, b)\mu(n, c)\xi(b, c, d)\eta(m, n, p))$$

is row finite.

Now consider the vector space V over Φ of all mappings on the product set $B \times M \to \Phi$. If $b \otimes m$ is the mapping which assigns 1 to the pair (b, m) and zero to everything else, then every element x of V can be expressed uniquely in the form

(3) $$x = \Sigma\lambda(m, b)b \otimes m.$$

The scalar multiplication, of course, is given by

(4) $$\alpha x = \Sigma\alpha\lambda(m, b)b \otimes m, \qquad \alpha \in \Phi.$$

Consider the subspace W of V consisting of all vectors (3) for which $(\lambda(m, b))$ is a row finite matrix. If we define a multiplication in W by means of (2), then with respect to (2) and (4), W is an algebra over Φ.

Let us assume that some b is the unit element 1 of S. Then the subalgebra of W consisting of all vectors $\Sigma\lambda(1, m)1 \otimes m$, where $\lambda(1, m)$ is always m-finitely valued or m-infinitely valued depending upon whether R is a polynomial or a power series algebra, is isomorphic to $\Phi[X]$ or $\Phi\{X\}$ respectively, and we shall denote it by $1 \otimes \Phi[X]$ (resp. $1 \otimes \Phi\{X\}$.) Similarly the set of vectors $\Sigma\lambda(b, 1)b \otimes 1$, where $1 = X_1{}^0X_2{}^0 \cdots X_n{}^0$, and where $\lambda(b, 1)$ is always b-finitely valued, forms a subalgebra of W isomorphic to S, which we shall denote by $S \otimes 1$. The algebra W^* generated by *finite* sums of products of elements from $1 \otimes \Phi\{X\}$ and $S \otimes 1$ does not equal W unless S is finite dimensional, but in the polynomial case $W^* = W$ without any restriction on S.

In the polynomial case W is simply the Kronecker product of the algebras S and $\Phi[X]$, but in the power series case, when $(S:\Phi) < \infty$, W is some sort of a generalized Kronecker product. Let $D = \Phi[X]$ or $\Phi\{X\}$ depending

upon whether $R = S[X]$ or $S\{X\}$. Then we shall write $W = S \otimes D$, and call it the Kronecker product [12] of S and D.

From (1), (2), (3), and (4) we see that the mapping

$$\Sigma\lambda(b, m)bm \to \Sigma\lambda(b, m)b \otimes m \tag{5}$$

is an isomorphism of R onto $S \otimes D$. We require now the following result.

THEOREM. *If U is an ideal in D, then $S \otimes U$ is an ideal in $S \otimes D$, and the mapping $U \to S \otimes U$ is a lattice isomorphism of the lattice of ideals of D onto the lattice of ideals of $S \otimes D$, provided either $(S:\Phi) < \infty$ or $D = \Phi[X]$.*

In the polynomial case, where $S \otimes D$ is the usual Kronecker product, the result was proved by Nakayama and Azumaya [1]. When $(S:\Phi) < \infty$, the following observation and method used by Jacobson [13] in proving the theorem of Nakayama and Azumaya, leads to a proof for the power series case. The essential point in Jacobson's proof is that if $v \in S \otimes D$, then v can be expressed as a finite sum $v = \sum_{i=1}^{n} a_i \otimes f_i$, $a_i \in S$, $f_i \in D$, where the a_i are linearly independent in S. From our remarks above, this can be done if and only if $(S:\Phi) < \infty$.

If the conditions of the theorem hold, then it is easy to verify that if A and B are ideals in D, then

$$S \otimes AB = (S \otimes A)(S \otimes B). \tag{6}$$

From the Theorem and (6) it follows that if Q is a primary ideal in D, then $S \otimes Q$ is primary in $S \otimes D$. Since D is a Noetherian ring, the theorem impiles that $S \otimes D$ has a Noetherian ideal theory, and finally, from (5), since R is isomorphic to $S \otimes D$, we conclude that R has a Noetherian ideal theory.

UNIVERSITY OF WISCONSIN.

---

[12] It is not difficult to show that the structure of W is independent of the particular bases we have chosen for S and D.

[13] Cf. Theorem 3, Chapter VI, of a book on the structure theory of rings, to appear in the Annals of Mathematics Studies.

REFERENCES.

[1] G. Azumaya and T. Nakayama, "On irreducible rings," *Annals of Mathematics*, vol. 48 (1947), pp. 949-965.

[2] R. Baer, "Radical ideals," *American Journal of Mathematics*, vol. 65 (1943), pp. 537-568.

[3] R. Dilworth, "Non-commutative residuated lattices," *Transactions of the American Mathematical Society*, vol. 46 (1939), pp. 426-444.

[4] H. Fitting, "Primärkomponentenzerlegung in nichtkommutativen Ringe," *Mathematische Annalen*, vol. 111 (1935), pp. 19-41.

[5] L. Fuchs, "On primal ideals," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 1-8.

[6] L. Fuchs, "On a special property of the principal components of an ideal," *Det Kongelige Norske Videnskabers Selskab*, vol. XXII (1950), pp. 28-30.

[7] N. Jacobson, "*The theory of rings*," Mathematical Surveys, Number 2, 1943.

[8] ———, "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

[9] W. Krull, "Ein neuer Beweis für die Hauptsätze der allgemeinen Idealtheorie," *Mathematische Annalen*, vol. 90 (1923), pp. 55.

[10] ———, "Zweiseitige Ideale in nichtkommutativen Bereichen," *Mathematische Zeitschrift*, vol. 28 (1928), pp. 481-503.

[11] ———, "Idealtheorie in Ringen ohne Endlichkeitsbedingung," *Mathematische Annalen*, vol. 101 (1929), pp. 729-744.

[12] ———, "Dimensiontheorie in den Stellenringen," *Journal für die Reine und Angewandte Mathematik*, vol. 179 (1938), pp. 204-226.

[13] J. Levitzki, "On multiplicative systems," *Compositio Mathematica*, vol. 8 (1950), pp. 76-80.

[14] ———, "Prime ideals and the lower radical," to appear in the *American Journal of Mathematics*.

[15] N. H. McCoy, "Prime ideals in general rings," *American Journal of Mathematics*, vol. 71 (1948), pp. 823-833.

[16] D. Murdoch, "Intersections of primary ideals in a non-commutative ring," *Bulletin of the American Mathematical Society*, vol. 56 (1950), Abstract 456.

[17] E. Noether, "Idealtheorie in Ringbereichen," *Mathematische Annalen*, vol. 83 (1921), pp. 24-66.

TWO DECOMPOSITION THEOREMS FOR A CLASS OF FINITE ORIENTED GRAPHS.*

By R. Duncan Luce.

1. Introduction.[1] The object of study in this paper is the class of finite oriented graphs which are subject to the conditions:

i. at most two branches exist between any pair of nodes (vertices), and

ii. whenever two branches do exist between a pair of nodes they shall have the opposite orientation.

Such a system will be called a network. The justification for introducing this word is its wide use in those applied sciences where oriented graphs of this type are playing an important role; for example: electrical networks, sociometric networks or diagrams, abstract programs for digital computers, and the neural networks of mathematical biology.

It is convenient to give a self-contained definition: A *network* N of order m is a system composed of two sets M and P, M being a finite non-empty set of m elements called the *nodes* of N and P a prescribed subset of the set of all ordered pairs of nodes. The members of P (i. e. the oriented branches) are called the *links* of N. The number of links of a network N will be denoted by $p(N)$, or simply by p when no ambiguity can arise. To indicate that N is of order m and has p links we shall write $N = N(m, p)$. Lower case Latin letters such as $a, b, c, \cdots$ will be used for nodes, and bracketed ordered pairs $(ab), (ca), \cdots$ to denote links. If (ab) is a link, the first node, a, will be called the *initial node* and the second, b, the *end node* of the link.

* Received October 2, 1951.

[1] Several of the concepts defined in the introduction have been assigned terms by D. König, *Theorie der Endlichen und Unendlichen Graphen*, New York, Chelsea Publishing Co., 1950. A brief glossary with page references to König is presented:

node	= Knotenpunkt, p. 1
link	= Gerichtete Kante, p. 4
disjoint	= Fremd, p. 3
arc of a network	= Zweifache Kante, p. 93
link of the form (aa)	= Schlinge, p. 3
non-reflexive graph	= Graph im engeren Sinn, p. 4
circuit	= Zyklus, p. 29
chain which is not a circuit	= Bahn, p. 30.

A *subnetwork* N' of a network N is a subset M' of the nodes, M, of N, with P' taken to be some subset (not necessarily proper) of those links of N which are definable on M'. If $M' = M$, we shall say the subnetwork is *complete*. Two subnetworks of a given network are *disjoint* if they have no nodes, and therefore no links, in common.

Each network is obviously a binary relation over a finite set, its nodes, and conversely every binary relation over a finite set can be interpreted as a network. This allows us to present all examples as relation matrices with entries 0 and 1 from the two element Boolean algebra. Furthermore, this suggests that if N and N' are two networks over the same (or isomorphic) set of nodes M, then by $N - N'$ we shall mean the complete subnetwork of N having those links of N which are not links of N'. If N' is a subnetwork of N, and if N' has the set of links P', then by the network formed from N by the *removal* of the links P', we mean $N - N'$. If N' has but one link, (ab), of N, then we shall write $N - N' = N - (ab)$.

We shall call a network *non-reflexive* if there are no links of the form (aa).

In case both the links (ab) and (ba) are present in a network, we shall say that an *arc ab* exists between a and b, the arc consisting of this pair of links, each of which will be said to be a member of the arc. This terminology is justified by the fact that when every link is a member of an arc the network is isomorphic (in the obvious sense of the word) to a graph without 2-circuits, to use a term of Whitney [2]; this is what we shall mean by saying that *a network is a graph*. Observe that the arcs of a network N are not the same as the branches (or arcs) of the graph which is oriented to form N. A link of the form (aa) is always the arc aa.

A (connected and oriented) *q-chain* from a to b is a set of q links of the form $(ac_1), (c_1c_2), \cdots, (c_{q-2}c_{q-1}), (c_{q-1}b)$, such that no node appears more than once, except in the case $a = b$ where a appears twice. Any q-chain from a to b will be denoted by (ab, q). Observe that $(ab, 1) = (ab)$. If c is a node included in a q-chain from a to b, then we may subdivide the chain into the "product" of two chains, one from a to c, and the other from c to b, i. e., $(ab, q) = (ac, q')(cb, q - q')$, $q' < q$.

An (oriented) *circuit* is a chain of the form (aa, q). A circuit of two links is an arc and conversely.

A network is *connected* if there exists a chain from each node to every other node. A network which is not connected is *disconnected*. When N is

---

[2] Whitney, H., "Non-separable and planar graphs," *Transactions of the American Mathematical Society*, vol. 34 (1932), p. 339.

treated as an oriented graph, connectedness is defined topologically; our definition implies topological connectedness but is not implied by it. However, as applied to networks which are graphs, the two definitions are equivalent.

2. A decomposition theorem for arbitrary networks. In this section we shall give four related definitions for networks of order m which will depend on the integers k between 0 and m. These definitions will be used in their full generality in Theorem 2.4, which shows that, in a certain sense, we need only consider the definitions in the case $k = 1$. Consequently the rest of the paper will be, for the most part, devoted to that special case.

First we need a measure of how easily a connected network is disconnected by the removal of links. We shall say a network is of degree 0 if it is not connected. A network is of *degree* k, $1 \leq k \leq m$, if there exists a set of k distinct links whose removal from the network will result in a complete subnetwork of degree 0, while the removal of any set of $q < k$ links results in a complete connected subnetwork.[3] The degree of a network is unique.

LEMMA 2.1. *If $N(m, p)$ is a network of degree k, then $p \geq km$.*

Proof. It will obviously suffice to show that each node is the initial node of at least k links. This is true, for if not, then the removal of the links for which such a node is the initial node will disconnect N. This contradicts the assumption that the degree is k.

In addition to the concept of degree, we need a condition implying that there is an even distribution of connectedness throughout the network; roughly, that the degree of any connected subnetwork is not greater than that of the network itself. That this is not always the case is evidenced by any graph formed of an $m - 2$ simplex, $m \geq 4$, and a single node joined by a single arc to one of the nodes of the simplex. The network is of degree 1, and the simplex, which is a connected subnetwork, is of degree $m - 2 \geq 2$. A definition which will suffice is the following. A network is said to be *k-minimal*, $1 \leq k \leq m$, if the removal of any link results in a complete subnetwork of degree $k - 1$. The existence of such networks is proved in Lemma 2.3. A network is *k-uniform* if every connected subnetwork is of degree $\leq k$. If a network is 1-uniform and connected we say it is *uniform*.

LEMMA 2.2. *If N is a k-minimal network with $k \geq 2$, then N is k-uniform and of degree k.*

---

[3] This definition of degree has no relation to the *Grad* defined by König, *op. cit.*, p. 3.

Proof. Let S be any connected subnetwork of N, and suppose it has degree d. Let $(ab) \varepsilon S$. $N - (ab) = N'$ is of degree $k - 1$, so in N' there exists a set U of $k - 1 \geqq 1$ links, whose removal from N' results in a complete disconnected subnetwork, N''. If in N'' there is a chain from a to b, we may replace (ab) and still have a disconnected network N^* which is formed from N by removing the links of U. In that case, the removal of any $(cd) \varepsilon U$ from N results in a complete subnetwork of degree $k - 2 \geqq 0$, since $k \geqq 2$. This is contrary to the assumption that N is k-minimal, so there is no chain from a to b. Thus, the removal of no more than k links, those of U which are in S and (ab), from S, implies a is not connected to b by any chain. It follows that $d \leqq k$.

Specifically, N has degree $d \leqq k$. If $d \leqq k - 1$, then, since the removal of any link results in a complete subnetwork of degree $k - 1$, it follows that $d = k - 1$. Let U be a set of $k - 1$ links whose removal from N results in a complete disconnected subnetwork. U is non-empty since $k \geqq 2$. Remove $(ab) \varepsilon U$ *from* N. The resulting network is, by definition, of degree $k - 1$; however, the remaining $k - 2$ links of U disconnect $N - (ab)$. Hence $d = k$.

We note that the above argument does not apply for $k = 1$; in fact, any disconnected network is 1-minimal, since the removal of any link results in a complete subnetwork of degree 0. But some networks are both connected and 1-minimal; these we shall call *minimal*. A minimal network is clearly non-reflexive and uniform.

LEMMA 2.3. *If N is a network of degree $k \geqq 1$, then for every integer q, $1 \leqq q \leqq k$, there exists a complete connected subnetwork of N which is q-minimal.*

Proof. Let C_q be the set of all complete connected subnetworks of N having degree q. Since N is finite and $q \leqq k$, it is obvious that C_q is non-empty. Let

$$p_q = \max_{N' \varepsilon C_q} p(N - N').$$

Since N is finite, there exists some $N_q \varepsilon C_q$ such that $p(N - N_q) = p_q$. N_q is, by choice, connected. Hence it will suffice to show that N_q is q-minimal. Suppose the removal of some link does not result in a complete subnetwork of degree $q - 1$. Then, since the removal of one link cannot lower the degree by more than 1, the resulting network N' has degree q. Thus $N' \varepsilon C_q$ and $p(N - N_q) < p(N - N') \leqq p_q$, which is contrary to choice.

A complete connected subnetwork N' of N such that, in the terms of the

above proof, $p(N - N') = p_q$, is called a *q-descendant* of N. If N_q and N'_q are two q-descendants of a network N, then $p(N_q) = p(N'_q)$. It is clear that every connected network has at least one 1-descendant, but this is not generally true for $q > 1$. Because of their importance the 1-descendants will be called simply *descendants*. It is clear that a descendant is minimal.

A network N will be called the sum of complete subnetworks N_i, $i = 1, 2, \cdots, t$, and written $N = \sum_{i=1}^{t} N_i$, if each link of N is contained in exactly one of the N_i.

THEOREM 2.4 (*first decomposition theorem*). *To every network N there exists a unique number k, its degree, and at least one set of $k+1$ complete 1-minimal subnetworks, N_i, such that*

i. $N = \sum_{i=1}^{k+1} N_i$,

ii. *N_{k+1} is disconnected,*

iii. *if $k \geq 1$, then N_1 is minimal,*

iv. *$\sum_{i=1}^{j} N_i$ is a j-descendant of $\sum_{i=1}^{j+1} N_i$, $1 \leq j \leq k$,*

and

v. *the connected subnetworks of the N_i, $1 \leq i \leq k$, are minimal, and so these networks N_i are 1-uniform.*

Proof. By definition there is a unique degree k assigned to every network. If $k = 0$, then N is not connected and we are done. If $k > 0$, select, according to Lemma 2.3, a k-descendant N'_k of N, and define $N_{k+1} = N - N'_k$. N_{k+1} is not connected; for if so, N is the sum of two complete subnetworks having, respectively, degree k (Lemma 2.2) and degree ≥ 1. This, we will show, implies that N has degree $\geq k+1$, which is contrary to assumption.

To show this we prove the slightly more general statement: If $N = N_1 + N_2$, and these networks have degrees k, k_1, and k_2 respectively, then $k \geq k_1 + k_2$. For, by definition, there exists a set U having k links, such that their removal from N results in a complete disconnected subnetwork N', and this is not true for any smaller set. Of these k links, let u_1 be in N_1, and u_2 in N_2. By the definition of a sum, $k = u_1 + u_2$. Furthermore, $u_1 \geq k_1$, since we may remove from N first the links of U and then the remaining links of N_2. This complete subnetwork, which obviously is N_1 minus u_1 links, is disconnected, so $u_1 \geq k_1$. Similarly, $u_2 \geq k_2$, whence the result.

In the network N'_k select a $(k-1)$-descendant, N'_{k-1}, and let $N_k = N'_k$

$-N'_{k-1}$. N_k is 1-minimal, for if $(ab) \varepsilon N_k$, let $N^*_k = N'_k - (ab)$. Then, by the definition of k-minimal, N^*_k is of degree $k-1$. But since N^*_k contains N'_{k-1}, the latter is a $(k-1)$-descendant of the former. Then the argument given above shows that $N^*_k - N'_{k-1} = N_k - (ab)$ is not connected.

The argument proceeds inductively without difficulty, since the last argument is independent of k. When we get to the case N'_2, $N'_1 = N_1$ is a descendant of N'_2 and thus is minimal rather than simply 1-minimal.

Condition (iv) is satisfied by our choice of the N_i.

Finally, the connected subnetworks of N_j, $2 \leq j \leq k$ are minimal. For suppose S is a connected subnetwork of N_j such that $S - (ab)$ is a connected subnetwork of S. Let $N^*_j = N'_j - (ab)$. Since N'_j is j-minimal, N^*_j is of degree $j-1$, so there exists a set of $j-1$ links whose removal from N^*_j will result in a complete disconnected subnetwork. At least one of these links is in S, since there exists, in S, a chain $(ab, q) \neq (ab)$. Thus there are at most $j-2 \geq 0$ of these links not in N_j, so that the removal of at most $j-2$ links from the descendant N'_{j-1} results in a complete disconnected subnetwork. This is in contradiction to Lemma 2. 3 which shows that N'_{j-1} is $(j-1)$-minimal. Thus S is minimal. If $k \geq 1$, N_1 is minimal, and therefore the connected subnetworks are minimal. It follows immediately that these N_i are 1-uniform.

In the sense of this theorem, the study of an arbitrary network has been reduced to the study of a collection of 1-minimal networks. These 1-minimal networks are either connected, and so minimal, or disconnected. But a disconnected network consists of isolated nodes, isolated chains, and connected pieces. For k of the subnetworks, part (v) shows that the connected pieces are minimal. If the theorem is applied repeatedly to the connected pieces of N_{k+1}, it may, in the same sense, be reduced to isolated nodes, isolated chains, and minimal subnetworks. Thus we may say that, in a sense, the study of any network may be reduced to the study of minimal networks. This exaggerates the present state of the art, since we do not know whether this decomposition is sufficiently strong to allow general conclusions about networks, or even k-minimal networks, from a knowledge of minimal networks. In fact, an important unsolved problem is the relationship between two distinct decompositions of this type for a given network. That two distinct decompositions may exist is shown by:

$$\begin{bmatrix} 01000 \\ 00101 \\ 01010 \\ 01100 \\ 10010 \end{bmatrix} = \begin{bmatrix} 00000 \\ 00000 \\ 01000 \\ 00100 \\ 00010 \end{bmatrix} + \begin{bmatrix} 01000 \\ 00101 \\ 00010 \\ 01000 \\ 10000 \end{bmatrix} = \begin{bmatrix} 00000 \\ 00100 \\ 00010 \\ 01000 \\ 00000 \end{bmatrix} + \begin{bmatrix} 01000 \\ 00001 \\ 01000 \\ 00100 \\ 10010 \end{bmatrix}.$$

On the basis of the preceding remarks we are led to devote the rest of this paper to beginning a study of minimal networks. Section 3 includes a decomposition of any minimal network and the deduction of several properties of minimal networks. These properties are used in section 4 to draw some conclusions about arbitrary connected networks. In section 5 we discuss the relationships between several of our concepts and that of a tree in graph theory. Finally, in the last section, we present an interesting inequality, and, from this, define a subclass of minimal networks, the members of which are shown to have a particularly simple form.

3. A decomposition theorem for minimal networks. This section presents a decomposition theorem for any minimal network, which may be used to show that there exists a close connection between the concept of a minimal network and the concept of a tree in graph theory. We note first that a network which is a tree is minimal. However, the class of minimal networks is much wider than that, for we know that every connected network has a descendant, which is minimal, and we have

LEMMA 3.1. *If N is a connected network and T a descendant of N, then T is a tree only if $T = N$.*

Proof. If T is a tree and $T \neq N$, then T must have been formed from N by the removal of at least one link. Reintroduce one of these, say (ab), into T. This must introduce an oriented circuit on at least three nodes, since T is connected; let it be $(ab)(bc_1)\cdots(c_qa)$. Since this circuit is on three or more nodes, and the links of T are members of arcs, it follows that $(ab, q) = (ac_q)(c_qc_{q-1})\cdots(c_1b)$, $q \geq 2$, exists. Because of the existence of the circuit, this chain may be removed to result in the complete connected subnetwork N'. Now since $q \geq 2$, $p(N - N') > p(N - T)$, so T is not a descendant of N, which is contrary to assumption.

To carry further the work of this section we need two more definitions. First, a network is a *compound circuit of order* 1 if it is simply a non-reflexive oriented circuit on its nodes; assuming a compound circuit of order $s - 1$ defined, a *compound circuit of order s* is formed from one of order $s - 1$ by replacing some node c of that network by a non-reflexive circuit C, each link of the form (ac) by one and only one link of the form (ac') where $c' \in C$, and each link of the form (ca) by one and only one link of the form $(c''a)$, $c'' \in C$. We shall refer to this as an *inductive composition* of a compound circuit. Obviously any compound circuit is connected; furthermore, we have

LEMMA 3.2. *If $N(m,p)$ is a compound circuit of order s, then $s=p-m+1$; i.e., if N is formed by orienting the graph G, s is the first Betti number of G.*[4]

Proof. Let N be formed by the inductive composition of the circuits C_i, $i=1,2,\cdots,s$, in their natural order, and suppose each C_i has m_i nodes and hence the same number of links. It follows by a simple induction that

$$m=m_1-1+m_2-1+\cdots+m_s=\Sigma m_i-(s-1)$$

and $p=\Sigma p_i=\Sigma m_i$, so that $s=p-m+1$. It is well known that $p-m+1$ is the first Betti number of any connected graph having p branches and m nodes.

A connected network N will be said to be *reducible* into subnetworks N_1 and N_2 if:

i. N_1 and N_2 are disjoint,

ii. N_1 and N_2 are each either connected or consist of a single node,

iii. there exists a network N', formed of N_1 and N_2 joined by exactly one link from N_1 to N_2 and exactly one from N_2 to N_1, such that $N'=N$.

If a connected network is not reducible it is called *irreducible.*

THEOREM 3.3. *A connected network which is a graph is reducible if and only if it is of degree* 1.

Proof. It is clear that any reducible network is of degree 1, since we may disconnect it by removing either of the links joining the disjoint subnetworks.

Let N be a graph of degree 1 and (ab) a link such that $N-(ab)=N'$ is disconnected. Evidently, in N' there is no chain from a to b. Define M_b to consist of b and any nodes b' such that there is a chain from b' to b in N'. Let $M_a=M-M_b$. Clearly, $a\varepsilon M_a$. For any $a'\varepsilon M_a$, $a'\neq a$, there exists in N' a chain from a to a'. If not, then, since N is connected, any chain in N from a to a' must contain the link (ab), and at least one such chain exists. Since the node a can appear only one, this chain must be of the form $(aa',q)=(ab)(ba',q-1)$, and $(ba',q-1)$ does not contain a. But N is a graph, so $(ba',q-1)$ implies the existence of a chain $(a'b,q-1)$ which does not contain (ab). This, then, is a chain in N', and so $a'\varepsilon M_b$, which is

---

[4] König, *ibid.*, first Betti number = Zusammenhangszahl, p. 53; Whitney, *op. cit.*, first Betti number = nullity, p. 340.

contrary to choice. Thus we know a chain exists in N' from a to a'. Since N is a graph, this implies that the largest subnetworks of N defined on M_a and M_b are each either connected or consist of a single node. Now, if $a' \varepsilon M_a$ and $b' \varepsilon M_b$, where $a' \neq a$ or $b' \neq b$, then there exists no link of the form $(a'b')$, for otherwise there would be a chain from a to b in N'. Since N is a graph, it follows that there are no links of the form $(b'a')$. Thus N is reducible.

The following is our principal result:

THEOREM 3.4 (*second decomposition theorem*). *To any minimal network N, which is not a tree, there exist integers $t \geq 1$ and $y \geq 0$, such that N consists of t disjoint irreducible compound circuits C_i, $i = 1, 2, \cdots, t$, and y nodes C_{i+t}, $i = 1, 2, \cdots, y$, not included in the C_i, $1 \leq i \leq t$, subject to the conditions:*

i. *there exists at most one link from any C_i to any C_j, $i \neq j$, $1 \leq i$, $j \leq t + y$;*

ii. *no arc is contained in any of the C_i, $1 \leq i \leq t$;*

iii. *the network formed by treating the C_i, $1 \leq i \leq t$, as nodes, all other nodes and links remaining unchanged, is a tree, or, if $t = 1$ and $y = 0$, a single node.*

Proof. This proof will be carried out in two stages. First we shall show that if N is a minimal network in which there exists an arc ab, then N is reducible into two subnetworks joined only by ab. Since N is minimal, it is non-reflexive, so that $a \neq b$. Define the set of nodes M_a to consist of a and any other nodes, a', of N such that there exists a chain from a to a' which does not include the link (ab). Let $M_b = M - M_a$. $b \varepsilon M_b$, for if not, then $b \varepsilon M_a$, and so there exists (ab, q) not including (ab). Then $N - (ab)$ is a connected subnetwork of N; this violates the condition that N is minimal.

We shall now show some properties N must satisfy which will lead ultimately to a proof of the statement:

1. If $a' \varepsilon M_a$, $a' \neq a$, there exists a chain from a' to a not including the link (ba). Clearly some chain exists from a' to a, since N is connected. If all such chains include (ba), then, since the node a may only appear once, each of them may be written in the form $(a'a, q) = (a'b, q-1)(ba)$. Moreover, $(a'b, q-1)$ does not include (ab) since $a' \neq a$. Now, by the definition of M_a, there exists a chain (aa', u) which does not include (ab), so that $(aa', u)(a'b, q-1)$ does not include (ab). This is contrary to the assumption that N is minimal.

2. Let $b' \neq b$. $b' \varepsilon M_b$ if and only if there exists a chain from b to b' which does not include the link (ba). Suppose first that $b' \varepsilon M_b$. Since N is connected, there exists at least one chain from b to b'. Suppose each (bb', q) contains (ba). Then, since each node may appear only once, $(bb', q) = (ba)(ab', q-1)$. If $(ab', q-1)$ does not contain (ab), then, by definition, $b' \varepsilon M_a$, which is impossible. But $(ab', q-1)$ cannot contain (ab), for if it did, then (bb', q) would not be a chain. Hence (bb', q) does not contain (ba).

Conversely, suppose there exists a chain from b to b' not including (ba). If $b' \varepsilon M_a$, there exists, by 1, a chain from b' to a not including (ba); these two combine into a chain from b to a not including (ba), which is contrary to N being minimal.

3. If $b' \varepsilon M_b$, $b' \neq b$, there exists $(b'b, q)$ not including (ab). We may parallel the proof of property 1 by replacing the words "definition of M_a" by "property 2."

4. If $a' \varepsilon M_a$, $b' \varepsilon M_b$, and either $a \neq a'$ or $b \neq b'$, then no link of the form $(a'b')$ exists. Suppose such a link does indeed exist. Then, by the definition of M_a, there exists a chain (aa', q) which does not include (ab), and, by property 3, a chain $(b'b, r)$ which does not include (ab). Thus the chain $(aa', q)(a'b')(b'b, r)$ does not include (ab), since $(a'b') \neq (ab)$, which is impossible.

5. Under the same conditions as in 4, there is no link of the form $(b'a')$. The argument is exactly parallel to that of 4, using properties 1 and 2.

It thus follows that the maximal subnetworks of N on the sets M_a and M_b are each either connected and minimal, or consist of a single node. From 4 and 5, one concludes that the subnetworks are joined only by the arc ab. This exhausts N, and the result is proved.

Since an arbitrary network has a finite number of arcs, it follows from a finite number of applications of the above result that a minimal network which is not a tree consists of a set of $t' \geq 1$ disjoint arc-free minimal subnetworks C_i, $i = 1, 2, \cdots, t'$, and $y' \geq 0$ nodes $C_{t'+i}$, $i = 1, 2, \cdots, y'$, not included in the C_i, $1 \leq i \leq t'$, such that:

i. any link not in a C_i, $1 \leq i \leq t'$, is a member of an arc;

ii. there exists at most one arc between any C_i and C_j, $i \neq j$, $1 \leq i$, $j \leq t' + y'$;

iii. the network formed by treating the C_i, $1 \leq i \leq t'$, as nodes, is a tree;

iv. the decomposition is unique.

By virtue of this decomposition, the problem is reduced to examining the case of an arc-free minimal network. We show: An arc-free minimal network consists of $t'' \geq 1$ disjoint irreducible compound circuits C_i, $i = 1, 2, \cdots, t''$, and $y'' \geq 0$ nodes $C_{i+t''}$, $i = 1, 2, \cdots, y''$, not included in the C_i, $1 \leq i \leq t''$, such that:

i. there exists at most one link from any C_i to any C_j, $i \neq j$, $1 \leq i$, $j \leq t'' + y''$;

ii. the network formed by treating the C_i, $1 \leq i \leq t''$, as nodes is a tree, or, if $t'' = 1$ and $y'' = 0$, a single node.

The first arc-free minimal network occurs for $m = 3$, and this obviously satisfies the conditions since it is a circuit on three nodes. Suppose now that the statement, except for the condition that the compound circuits are irreducible, has been proved for all networks through $m - 1$ nodes, and let $N(m, p)$ be an arc-free minimal network. In N there exists a circuit consisting of at least three links, since N is connected, arc-free, and non-reflexive; let C be one such circuit on the nodes c_i, $i = 1, 2, \cdots, q \geq 3$. The maximum subnetwork of N on these nodes is only C, for if there exists any other link $(c_i c_k)$, $k \neq i + 1$, then this link can be removed without disconnecting N, since the chain $(c_i c_{i+1})(c_{i+1} c_{i+2}) \cdots (c_{k-1} c_k)$ exists. This is impossible since N is minimal. Now, if C exhausts all the nodes of N we are done. If not, let M' be the set of nodes remaining. If $a \in M'$, and (ac_k), $1 \leq k \leq q$, exists, then no link of the form (ac_i), $1 \leq i \leq q$, $i \neq k$, exists. For if so, then the chain from c_k to c_i, which is a part of C, and so does not include (ac_i), shows that (ac_i) may be removed without disconnecting N. This is impossible. Similarly, if $(c_k a)$, $1 \leq k \leq q$, exists, then $(c_i a)$, $1 \leq i \leq q$, $i \neq k$, does not exist.

Now consider the network N' formed by letting the nodes of C coalesce into a single node which we shall call c. Evidently, since N is minimal, so is N', and N' has at least two nodes fewer than N, since $q \geq 3$. Several possibilities exist for N'. First, it may be a graph, and hence a tree, in which case the statement is proved. Second, it is not a tree, but there exists at least one arc. By the first part of this theorem, N' may be decomposed into several arc-free minimal subnetworks connected in such a fashion that if they are treated as nodes, the resulting network is a tree. By the induction hypothesis, these arc-free minimal subnetworks satisfy the conditions of the statement we are proving. But replacing the node c by C, reconstructing N, only increases the order of one of these compound circuits, or introduces a new compound circuit, so the result is true for N. Third, N' is arc-free, in

which case the induction hypothesis may be applied directly, and the introduction of C for c only increases the order of the compound circuit.

Thus, we may decompose N into several compound circuits and nodes not in these compound circuits and connecting links satisfying the conditions i and ii of the second intermediate statement. Carry this decomposition as far as possible; the process will terminate in a finite number of steps, since N is finite. We will show that the resulting compound circuits are irreducible. For suppose that C_k is reducible into the disjoint subnetworks A and B connected by the links (ab), $(b'a')$, $a, a' \varepsilon A$, $b, b' \varepsilon B$. By condition ii it follows that any C_i, $i \neq k$, $1 \leq i \leq t' + y'$ is linked "symmetrically" to C_k if at all. In fact, it is either linked symmetrically to A or to B; for if not, then there exists a link from A to C_i and a link from C_i to B, in which case (ab) may be removed, or, in the other case, $(b'a')$ may be removed without disconnecting N. This is impossible. A and B are either compound circuits or, by the result proved for arc-free minimal networks, may be reduced to several compound circuits and nodes not in them such that i and ii hold. By an argument similar to the one just made, the conditions i and ii hold for N with this finer decomposition. This is contrary to choice, so C_k must be irreducible.

The proof of the theorem follows almost immediately from the two intermediate results, if we note that the last argument may be applied to show condition iii.

This decomposition of a minimal network is not unique, for

$$\begin{pmatrix} 010000 \\ 000001 \\ 010100 \\ 000010 \\ 001000 \\ 100010 \end{pmatrix}$$

may be decomposed into either a tree consisting of one arc, or one of two arcs.

The next result gives a little more information about the components into which we have decomposed a minimal network, the irreducible compound circuits. This result is unsatisfactory in the sense that it does not give a complete characterization of these networks. For this proof and succeeding results we need the following definition. A node is *simple* if it is the initial node of exactly one link and the end node of exactly one link.

THEOREM 3. 5. *Let N be a minimal network. N is irreducible if and*

only if it is a compound circuit such that in any inductive composition of N, none of the circuits introduced are arcs.

Proof. Suppose that at some stage of the composition of N, an arc ab is introduced into a compound circuit C to form a compound circuit C'. If ab does not have a simple node, then in C' there exists either a chain from a to b not containing (ab), or one from b to a not containing (ba), since such a chain exists in C. The introduction of further circuits can only lengthen this chain, so $N-(ab)$ is a complete connected subnetwork, which is impossible. Hence a node of ab is simple. The introduction of further circuits merely adds to C to form a larger compound circuit, and hence a connected subnetwork or a single node of N, or it may replace the simple node of ab by a compound circuit. Between these two connected subnetworks, or single nodes, are only the links arising from ab, now no longer an arc in general. Thus N is reducible, which is contrary to assumption, proving that no arc can be introduced.

Conversely, if we suppose N is reducible, then Theorem 3.4 implies N may be decomposed into one or more irreducible compound circuits and nodes not included in these compound circuits. The circuits of any compound circuit C may be coalesced into nodes in the inverse order of an inductive composition of C. This clearly leads to the tree of Theorem 3.4. But any tree is a compound circuit formed only of arcs. Thus we have an inductive composition of N involving arcs, the arcs of the tree. As this is contrary to assumption, N must be irreducible.

The principal theorem will be utilized sometimes through two properties of minimal networks derivable from it. They are presented as

THEOREM 3.6. *A minimal network is a compound circuit which contains at least two simple nodes.*

Proof. The last part of the above proof suffices to show that a minimal network is a compound circuit.

To show that a minimal network has two simple nodes, we shall perform an induction on the order s of the compound circuit. It is certainly true for $s=1$, since the compound circuit is then a non-reflexive circuit. Assume the result true up through circuits of order $s-1$. Suppose N is a minimal compound circuit of order s, and let C be the last circuit introduced in some inductive composition of N. Let C coalesce into a single node c, and call the resulting network N'. N' is readily seen to be minimal and of order $s-1$,

so by the induction hypothesis it contains at least two simple nodes. If two of these are different from c, then we are done. If not, c is simple. Consider N; if C is not an arc then it must introduce a simple node, for C has at least three nodes, and there exists only one link to C from the rest of the nodes, and only one from C, since c is simple. If, on the other hand, C is an arc, then the first argument in the proof of Theorem 3. 5 shows that one of its nodes is simple, and the result follows.

That not every compound circuit is minimal or has a simple node is shown by:

$$\begin{bmatrix} 0101 \\ 0010 \\ 0101 \\ 1000 \end{bmatrix}.$$

4. Applications to connected networks. Two applications to connected networks are given of the results on minimal networks; the first examines limits on the number of links a connected network may have, and the second discusses the maximum number of "independent" circuits a connected network may have.

THEOREM 4. 1. *Let $N(m, p)$ be a connected network, not a tree. Let N have a descendant N' which is decomposable in the terms of Theorem 3. 4 into t irreducible minimal subnetworks and y nodes not in these subnetworks. Then*

$$p \leq (3m + t + y - 4)/2 + p(N - N') < 2(m - 1) + p(N - N').$$

If N is a tree, $p = 2(m - 1)$.

Proof. If N is a tree, the result is well known from graph theory.[5]

Suppose N is not a tree. Then it is sufficient to show the result for the class of minimal networks which are not trees, since, in the general case, the network N has $p(N - N')$ more links than any descendant N'. By Lemma 2. 3 a descendant is minimal, and, by Lemma 3. 1, it is not a tree. So we consider N minimal. Decompose N as in Theorem 3. 4, and let the irreducible compound circuits C_i have m_i nodes, p_i links, and order s_i. Let there be p' links not in any irreducible compound circuit. Then, by the result on trees, $p' = 2(t + y - 1)$. For each of the irreducible minimal subnetworks, Theorem

[5] Whitney, *ibid.*, pp. 340-341.

3.5 implies that each of the s_i circuits used in forming C_i has at least three nodes, so that $m_i \geq 3 + 2 + 2 + \cdots + 2 = 2s_i + 1$. By Lemma 3.2,

$$p_i = s_i + m_i - 1 \leq 3(m_i - 1)/2.$$

Thus,

$$p = \Sigma p_i + p' \leq \Sigma 3(m_i - 1)/2 + 2(t + y - 1)$$
$$= (3/2)(\Sigma m_i + y) + (t + y - 4)/2 = (3m + t + y - 4)/2.$$

This may be simplified a little by noting that each of the irreducible minimal subnetworks must have at least three nodes, so $m \geq 3t + y$; hence,

$$p \leq (3m + t + y - 4)/2 = (4m - 4 - 2t + 3t + y - m)/2$$
$$\leq 2(m - 1) - t < 2(m - 1).$$

This concludes the proof.

It is clear that in a given network we may define the *addition of chains* (*mod* 2). Thus we may also define *linear independence* (*mod* 2). We shall be concerned with sets of linearly independent (mod 2) circuits such that no other linearly independent set contains a greater number of circuits. These sets will be called *maximal*. The result proved in the next theorem is, in statement, formally the same as a result of graph theory:[6]

THEOREM 4.2. *In any connected network $N(m, p)$ there exists a maximal set of $p - m + 1$ linearly independent (mod 2) circuits.*

Proof. First, it is sufficient to show this for minimal networks. For, if N is not minimal, then it has a descendant N' which is. N may be considered to be formed from N' by the addition of links one at a time. Each such link adds at least one new circuit which is independent (mod 2) of the circuits of the network to which it was added, since in a connected network every link is contained in at least one circuit. Thus, if there exists a set U' of $p - K - m + 1$, $K = p(N - N')$, linearly independent (mod 2) circuits in N', there will exists a set U of at least $p - m + 1$ linearly independent circuit in N.

Furthermore, if U' is maximal in the descendant, U will be in N also. If not, then there is a first subnetwork, N^*, for which any set of $p^* - m + 1$ linearly independent (mod 2) circuits is maximal, and to which the addition of a link (ab) produces a linearly independent set U'' having more than

[6] Lefschetz, Solomon, *Introduction to Topology*, Princeton, Princeton University Press (1949), p. 71.

$p^* - m$ circuits. It is clear that this set U'' must contain at least two circuits which include the link (ab), for otherwise the subset of U'' in N^* would contain more than $p^* - m + 1$ linearly independent circuits. Let two of the circuits be denoted by $(ab)(ba, q)$ and $(ab)(ba, q')$. Since N^* is connected and does not contain (ab), there exists a chain from a to b not including (ab); select a shortest: (ab, q''), $q'' > 1$. In general, (ab, q'') will coincide with (ba, q) over a certain number of links, i. e., over a set of several chains of the form (cd, t), each a part of (ba, q). The argument does not change in principal, and a great saving in notation is gained, if we assume that at most one such chain occurs. Similarly, (ab, q'') will be assumed to coincide with (ba, q') over the chain $(c'd', t')$. Furthermore, we shall assume that (cd, t) and $(c'd', t')$ have no links in common; if they do, a slight modification of the following argument will suffice. So we may write:

$$(ab, q'') = (ac, u)(cd, t)(dc', z)(c'd', t')(d'b, v),$$

the order of (cd, t) and $(c'd', t')$ being immaterial.

$$(ba, q) = (bc, x)(cd, t)(da, y), \qquad (ba, q') = (bc', x')(c'd', t')(d'a, y').$$

Observe that the following formal products are in fact circuits of N^*:

$$\text{A}: \quad (ac, u)(cd, t)(da, y) \qquad \text{B}: \quad (d'b, v)(bc', x')(c'd', t')$$

$$\text{C}: \quad (ac, u)(cd, t)(dc', z)(c'd', t')(d'a, y')$$

$$\text{D}: \quad (bc, x)(cd, t)(dc', z)(c'd', t')(d'b, v).$$

Since these are circuits of N^*, they are expressible (mod 2) in terms of the circuits in U''. But observe that $(ab)(ba, q) + A + B + C + D = (ab)(ba, q')$ (mod 2) so that $(ab)(ba, q)$ and $(ab)(ba, q')$ are not linearly independent. Thus only one of them can be in U'', and so we have shown that if the theorem is true for minimal networks it is true in general.

The minimal case will be proved by induction on m. For $m = 2$ it is trivially true. Suppose it is true for all minimal networks having $m - 1$ or fewer nodes, and let $N(m, p)$ be minimal. By Theorem 3. 6, N has a simple node a which is the initial node of only one link, (ab), and the end node of only one, (ca). We may distinguish three cases:

i. $b = c$. Remove a and the arc ab, leaving the subnetwork $N'(m - 1, p - 2)$. N' is obviously minimal, and so it has a maximal set of $p - m$ linearly independent (mod 2) circuits. The arc ab adds exactly one circuit to this set.

ii. $b \neq c$, and there does not exist $(cb, q) \neq (ca)(ab)$. Remove a and the adjoining links and introduce the link (cb) to form $N'(m-1, p-1)$, which is minimal. Thus, by the induction hypothesis, has a maximal set of $p-m+1$ linearly independent circuits. But forming N' from N cannot essentially change any set of linearly independent circuits, since the chain $(ca)(ab)$ is, in this case, formally the same as (cb).

iii. $b \neq c$, and there does exist $(cb, q) \neq (ca)(ab)$. Again remove a and the adjoining links to form $N'(m-1, p-2)$, which is minimal. By the induction hypothesis, N' has a maximal set of $p-m$ linearly independent circuits. Replacing a and the two links (ab) and (ca) to form N adds one or more new circuits, depending on the number of chains from b to c. This situation is not essentially different from the one discussed in the first part of this proof, except that we are adding a 2-chain and a new node, rather than a single link. Since this node is simple, the argument is formally the same, and it shows that there is a maximal set of $p-m+1$ linearly independent (mod 2) circuits in N. This, then, concludes the proof.

We note the trivial corollary: A connected network $N(m, p)$ has exactly $p-m+1$ circuits if and only if the set of all circuits of N is linearly independent (mod 2).

5. Generalizations of a tree. We shall show in this section that several of our definitions, when applied to networks which are graphs, are identical with the concept of a tree. This can be shown directly and easily in each case; however we shall first prove two results which are true in general, and then we shall use them to prove Theorem 5. 3. Thus, that result is not as deep as it first appears to be.

We shall call a connected network $N(m, p)$ having exactly $p-m+1$ circuits *circuit minimal.* This definition makes sense because of Theorem 4. 2. By the corollary to that theorem, N is circuit minimal if and only if N is connected and the set of all circuits is linearly independent (mod 2).

LEMMA 5. 1. *A circuit minimal network is uniform.*

Proof. Let $N(m, p)$ be circuit minimal. Let (ab) be any link, and $N-(ab)=N'$. If N' is not connected, then N has degree 1. If N' is connected it is circuit minimal, for at least one circuit of N was destroyed by the removal of (ab), and according to Theorem 4. 2, no more than one.

Since N' is connected, there exists at least one chain from b to a, but only one, for if there were more then the addition of the single link (ab) would introduce more than one circuit, and N would have more than $p - m + 1$ circuits. In N, interrupt the chain from b to a by removing a single link from it, thus disconnecting N. This proves N is of degree 1.

To show N is uniform it will thus suffice to show that every connected subnetwork is circuit minimal. If S is a connected network which is not circuit minimal, then there exists a circuit in S which is linearly dependent (mod 2) on the other circuits of S. This remains true in N, so N is not circuit minimal, a contradiction.

LEMMA 5.2. *A compound circuit is uniform.*

Proof. This may be demonstrated by an induction on the order of compound circuits. It is trivially true for compound circuits of order 1. Let N be a compound circuit of order $s > 1$, let C be the last circuit introduced in some inductive composition of N, and let S be any subnetwork of N. Coalesce C into a single node c, and under this operation let S become S'. If S' is a single node, then $S = C$, and the degree of S is 1. Otherwise, S' is connected, and therefore, by the induction hypothesis, it is of degree 1. Select $(ab) \varepsilon S, \not\varepsilon C$, such that $S' - (ab)$ is not connected. This is possible since S′ is of degree 1. Now the introduction in S' of that part of C in S, subject to the conditions of N, can only replace the node c by a chain or a circuit, but cannot introduce a link or a chain from a to b; thus S is of degree 1. So N is uniform.

The next result is the justification for the title of this section.

THEOREM 5.3. *For a connected network N which is a graph, the following are equivalent*: i. *N is a tree,* ii. *N is minimal,* iii. *N is a compound circuit,* iv. *N is uniform,* v. *N is circuit minimal.*

Proof. i. implies ii trivially. ii. implies iii by Theorem 3.6. iii. implies iv by Lemma 5.2. iv. implies i. For if N is not a tree, then there exists a circuit in the sense of graph theory. But this is clearly a connected subnetwork of degree 2, so that N is not uniform. v. implies iv by Lemma 5.1. i. implies v. If $N(m, p)$ is a tree, it follows, from theorem 3.5, that $p - (m - 1) = m - 1$. Furthermore, the only circuits in a tree are 2-circuits (arcs) of which there are exactly $m - 1$, so the number of circuits is $p - m + 1$.

The several results of this section and Theorem 3.6 suggest the following

class of unsolved problems: Conditions on a uniform network that it be a compound circuit. Conditions on a uniform network that it be circuit minimal. Conditions on a compound circuit that it be minimal. These four concepts are indeed all distinct. The network.

$$\begin{pmatrix} 010000 \\ 001001 \\ 000100 \\ 100010 \\ 010000 \\ 000100 \end{pmatrix}$$

is minimal, and hence a compound circuit, but not circuit minimal. The network

$$\begin{pmatrix} 0101 \\ 1000 \\ 1101 \\ 1010 \end{pmatrix}$$

is both uniform and circuit minimal, but not a compound circuit. The network

$$\begin{pmatrix} 0111 \\ 1010 \\ 1001 \\ 1000 \end{pmatrix}$$

is uniform, but neither a compound circuit nor circuit minimal. The example following Theorem 3. 6 shows that not every compound circuit is minimal.

6. Rank minimal networks. In section 1 we noted the representation of networks by relation matrices with entries from the two-element Boolean algebra. Equally well, we may interpret this as a representation by real matrices with the numbers 0 and 1 as entries. Thus, since it is well known that matrix rank is a similarity invariant, to each network there is a uniquely defined number r, $1 \leq r \leq m$, called the *rank of the network*, which is the rank of any of the corresponding real matrix representations.

THEOREM 6. 1. *If $N(m, p)$ is a connected network*[7] *having rank r, then* $p + r \geq 2m$.

---

[7] Luce, R. D., "Connectivity and generalized cliques in sociometric group structures," *Psychometrika*, vol. 15 (1950), pp. 169-190. In this paper the diameter, n, of a connected network was defined as $n = \max_{a,b \in M} \min_q (ab, q)$, and it was conjectured that $p + n \geq 2m$. This is now known to be false; however, Theorem 6. 1 is a correct result which is closely related to the conjecture, for it may also be shown that $r \geq n$.

Proof. Suppose $p + r < 2m$. Select any set R of r linearly independent rows in a particular matrix representation. Since N is connected, there exists a non-zero entry in each column j; but since each row can be written as a linear combination of rows from R, it follows that for each column j there exists an $i \varepsilon R$, such that the ij entry is 1. Thus, in the rows of R there are at least m 1's. By our assumption, there remain $p' \leq p - m < m - r$ links (entries that are 1). Each of the $m - r$ rows not in R must contain a non-zero entry, since N is connected, and therefore $p' \geq m - r$, a contradiction.

We shall call a network $N(m, p)$ *rank minimal* if it is connected, and $p + r = 2m$.

THEOREM 6.2. *If a connected network is rank minimal, it is minimal.*

Proof. As in the proof of Theorem 6.1, we consider a matrix representation N of the rank minimal network N, and let R be a set of r linearly independent rows. Each column has a non-zero entry in some row of R, since N is connected. The set R' of the $m - r$ remaining rows must have a non-zero entry in each row for the same reason. However, since $p = 2m - r$, it is necessary that R have exactly one non-zero entry in each column, and R' exactly one in each row.

Let (ab) be any link of the network. We shall show that its removal results in a disconnected network, which will prove the theorem.

If $a \varepsilon R'$, then the removal of (ab) results in a network N' having no link for which a is the initial node, since the rows of R' have exactly one non-zero entry.

If $a \varepsilon R$, then either the row a has only one non-zero entry, and we use the above argument, or it has another non-zero entry, say in column c, $c \neq b$. We show that in the latter case column b has only the one non-zero entry, N_{ab}. For suppose another link (db), $d \neq a$, exists. Then $d \varepsilon R'$, for we showed above, essentially, that the rows of R have exactly one non-zero entry in each column, and we have assumed (ab) to exist and $a \varepsilon R$. Since the rows of R' have exactly one non-zero entry, it follows that N_{db} is the one for row d. But since the rows of R are a set of linearly independent ones for this matrix, row d must be a linear combination of rows of R. The row a must be in this combination, as it is the only one of R having an entry in the b column. However, we assumed that row a has a non-zero entry in column c. This must be subtracted, since row d cannot have an entry $N_{dc} = 1$. But this is impossible using only rows of R, since no other row of R has an entry in the

c column. This contradiction implies that column b has only $N_{ab}=1$, and so $N-(ab)$ is a complete disconnected subnetwork of N. Thus N is minimal.

The converse statement is not true, as will be obvious from a comparison of Theorem 6.4 and Theorem 3.4.

The next lemma will be used in conjunction with Theorem 3.4 to decompose any rank minimal network.

LEMMA 6.3. *Let N be rank minimal. If N is reducible into the subnetworks N_1 and N_2, then either N_1 or N_2 is a single node.*

Proof. Let the N_i, $i=1,2$, have m_i nodes, p_i links, and rank r_i; and let m, p, and r denote the corresponding quantities in N. If neither of the N_i is a node, they are both connected subnetworks, so by Theorem 6.1, $p_i \geq 2m_i - r_i$. It is evident from the definition of a reducible network that $p=p_1+p_2+2$, and $m=m_1+m_2$. Furthermore, if we let the matrix minor repersentation of N_i be denoted by the same symbols, we then have, for an appropirate labeling of the nodes, the following type of matrix representation for N:

$$\begin{bmatrix} & 0\ \cdots\ 0\ \cdots\ 0 \\ & \vdots \qquad \vdots \qquad \vdots \\ N_1 & 0\ \cdots\ 1\ \cdots\ 0 \\ & \vdots \qquad \vdots \qquad \vdots \\ & 0\ \cdots\ 0\ \cdots\ 0 \\ 0\ \cdots\ 0\ \cdots\ 0 & \\ \vdots \qquad \vdots \qquad \vdots & \\ 0\ \cdots\ 1\ \cdots\ 0 & N_2 \\ \vdots \qquad \vdots \qquad \vdots & \\ 0\ \cdots\ 0\ \cdots\ 0 & \end{bmatrix},$$

whence one sees that $r \geq r_1 + r_2 - 1$. Thus, $p = p_1 + p_2 + 2 \geq 2m_1 - r_1 + 2m_2 - r_2 + 2 = 2(m_1+m_2) - (r_1+r_2-1) + 1 > 2m - r$, which is contrary to the assumption that N is rank minimal.

A tree such that the arcs all have one end node in common is called a star. It is not difficult to show that a star is rank minimal.

THEOREM 6.4. *Let N be a rank minimal network which is not a star. If N has any arcs, there exists one arc-free rank minimal subnetwork C of N, such that the network formed from N by treating C as a node, all other links and arcs remaining unchanged, is a star. An arc-free rank minimal network C consists of exactly one irreducible rank minimal subnetwork C', not a single node, and, possibly, some simple nodes such that the network formed from C by treating C' as a node, all other links and nodes remaining unchanged, is a star. Furthermore, if C' is the irreducible rank minimal subnetwork of the arc-free rank minimal subnetwork C of N, then the network formed from N by treating C' as a node, all other links and nodes remaining unchanged, is a star.*

Proof. In the first case, apply Lemma 6.3 to the first statement presented in the proof of Theorem 3.4 to show that each arc which exists must have a simple node. The arc-free subnetwork C is rank minimal, for the removal of any arc from N results in a network N' for which $p' = p - 2$, $m' = m - 1$, and $r' \leq r$, implying $p' \leq 2m' - r'$. Thus, by Theorem 6.1. N' is rank minimal. To C, first apply Theorem 3.4 and then Lemma 6.3 to show any nodes, not in an irreducible subnetwork, must be simple, and there is only one irreducible subnetwork C'. The same argument as applied above suffices to show that C' is rank minimal. The final statement is proved in the same manner.

In conclusion one may mention two more unsolved problems: Conditions on a minimal network that it be rank minimal, and a characterization of an irreducible rank minimal network.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

ON THE NON-VANISHING OF CERTAIN DIRICHLET SERIES.*

By Aurel Wintner.

Let $f(n)$, where $n = 1, 2, \cdots$, be a completely multiplicative function, that is, let $f(n_1n_2) = f(n_1)f(n_2)$ but $f(1) \neq 0$. Such a function is uniquely determined by an arbitrary assignment of the values $f(p)$, and is a bounded function if and only if $|f(p)| \leqq 1$ holds for every prime. The following theorem will be proved:

If $f(n)$ is completely multiplicative and bounded, and if the function $F(s)$, defined for $\sigma > 1$ by

$$F(s) = \sum_{n=1}^{\infty} f(n)/n^s, \tag{1}$$

has no singular point on $\sigma = 1$, then it has at most one zero on $\sigma = 1$.

This assertion is meant to imply that the zero, if any, cannot be a multiple zero. That it can occur at all, is shown by Liouville's example, $f(p) = -1$, where $s = 1$ is a zero of $F(s) = \zeta(2s)/\zeta(s)$. Since this $F(s)$ is replaced by $F(s - ia)$ if every $f(p) = -1$ is multiplied by p^{ia}, the zero can occur at any point, $s = 1 + ia$, of the line $\sigma = 1$.

For reasons of symmetry, the zero, if any, must be at $s = 1$ if $f(n)$ is real-valued. In this case, the assumption of boundedness, which is then equivalent to $-1 \leqq f(p) \leqq 1$, can be refined to $-1 \leqq f(p)$, if (1) is absolutely convergent for $\sigma > 1$. This was proved in [2] by an argument based on $\zeta(s)F(s)$. The above theorem will be proved by combining that argument with a device, introduced in this context by Ingham [1], which replaces $\zeta(s)F(s)$ by

$$G(s) = \zeta^2(s)F(s)F^*(s), \tag{2}$$

where $F^*(s)$ denotes the Dirichlet series the coefficients of which are the complex conjugates of the coefficients of (1).

First, if $\sigma > 1$, then, since $|f(n)| \leqq 1$, logarithmic differentiation of the Euler factorization of (1) gives

* Received June 15, 1951.

$$(3)\qquad -F'/F(s)=\sum_{n=1}^{\infty}\Lambda(n)f(n)/n^s,\text{ where }-\zeta'/\zeta(s)=\sum_{n=1}^{\infty}\Lambda(n)/n^s.$$

Hence, from (2),

$$(4)\qquad -G'/G(s)=\sum_{n=1}^{\infty}2\{1+\Re f(n)\}\Lambda(n)/n^s.$$

Since $F(s)$ is supposed to remain regular on $\sigma=1$, it is clear from (2) that the limit

$$(5)\qquad m=m(t)=\lim_{\epsilon\to 0}\epsilon G'/G(1+\epsilon+it)$$

exists for every real t and is the order of the zero $s=1+it$ of $G(s)$, with the understanding that this order can be negative or 0. On the other hand, since $|f(n)|\leqq 1$ and $\Lambda(n)\geqq 0$, every coefficient of (4) is non-negative. Hence it is clear from (5) that $|m(t)|\leqq|m(0)|$ holds for every t.

In particular $m(t)$ must vanish identically if it vanishes at $t=0$. This means that $G(s)$ must be regular and non-vanishing at every $s=1+it$ if it is regular and non-vanishing at $s=1$. But the latter assumption is satisfied if $F(s)$ vanishes at $s=1$ in the first order. This is clear from (2), since $\zeta(s)$ has a pole of first order at $s=1$. Consequently, if $F(s)$ has a simple zero at $s=1$, then $G(s)$ has no zero $s=1+it\neq 1$, which, in view of (2), means that $s=1$ is the only zero of $F(s)$ on the line $\sigma=1$. It follows that, in order to prove that

$$(6)\qquad F(1+it)\neq 0\text{ for every }t\neq 0\text{ if }F(1)=0,$$

it is sufficient to show that $F(s)$ cannot have a multiple zero at $s=1$.

Since $|f(n)\Lambda(n)|\leqq\Lambda(n)$, it is clear from (3) that

$$|F'/F(1+\epsilon)|\leqq|\zeta'/\zeta(1+\epsilon)|$$

if $\epsilon>0$. It follows therefore from

$$\lim_{\epsilon\to 0}\epsilon\zeta'/\zeta(1+\epsilon)=-1\text{ that }\lim_{\epsilon\to 0}|\epsilon F'/F(1+\epsilon)|\leqq 1.$$

Finally, the last inequality implies that $F(s)$ cannot have a multiple zero at $s=1$, i. e., that

$$(7)\qquad F'(1)\neq 0\text{ if }F(1)=0.$$

This proves (6). But (6) implies that

$$(8)\qquad F(1+it)\neq 0\text{ for every }t\neq a\text{ if }F(1+ia)=0,$$

where a is any real number. In fact, (8) follows if $f(n)$ in (1) is replaced by $f(n)n^{-ia}$ and then (6) is applied to the new function (1). Similarly, (7) implies that

$$F'(1+ia) \neq 0 \text{ if } F(1+ia) = 0. \tag{9}$$

Clearly, (8) and (9) together are equivalent to the theorem italicized above.

It is clear from the proof that, instead of assuming the regularity of $F(s)$ on $\sigma = 1$, it is sufficient to assume that $F(s)$ and $F'(s)$, where $\sigma > 1$, go over into continuous boundary values as $\sigma \to 1$. In fact, a somewhat less stringent condition would also suffice.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. E. Ingham, "Note on Riemann's zeta-function and Dirichlet's L-functions," *Journal of the London Mathematical Society*, vol. 5 (1930), pp. 107-112.

[2] A. Wintner, "On the non-vanishing of certain Dirichlet series," *Rendiconti del Circolo Matematico di Palermo*, New series, vol. 1 (1952).

ON THE FUNDAMENTAL GROUP OF AN ALGEBRAIC VARIETY.*

By Wei-Liang Chow.

It is well known that any 1-cycle in an algebraic surface can be deformed into a 1-cycle lying in a generic plane section of the surface.[1] The usual proof of this theorem, which can be easily generalized from a surface to any non-singular algebraic variety, is topological and consists of a simple construction of the deformation chain. In the transcendental theory there is a generalization of this theorem, at least in its homology aspect, which can be stated as follows:[2] There exist exactly $2p$ independent 1-cycles in an algebraic surface which are not homologous to 1-cycles belonging to a generic curve of an irrational pencil of genus p. In this paper we shall show that this theorem is a special case of a more general theorem about the fundamental group of an algebraic variety under a rational transformation. Our method of proof will be purely topological; the essential idea is that although a rational transformation is not in general a fibre mapping, the covering homotopy theorem is nevertheless true, in a somewhat modified form, for the mapping of a 1-simplex.

In section **1** the notion of a fibre system is introduced, and certain subsystems of an algebraic system are shown to be (or can be considered as) fibre systems. The notion of a fibre system is a generalization of that of a fibre space, and just as in the case of a fibre space we have also here as a fundamental property the validity of the covering homotopy theorem, which must now be formulated in a somewhat modified form. This notion of a fibre system is a very useful tool in the study of the topology of algebraic varieties; in this paper we shall limit ourselves strictly to the particular problem in question, but we hope to show in a later paper that the method is applicable also to other similar problems in algebraic geometry. In section **2** the results of section **1** will be used to prove two theorems; one of them (Theorem 2) is the theorem mentioned above, the other (Theorem 1) is a theorem concerning the deformation of 1-cycles into a member of an algebraic system with at least

* Received August 20, 1951.

[1] See, e.g., O. Zariski, *Algebraic Surfaces*, p. 108.

[2] See O. Zariski, *Algebraic Surfaces*, p. 144.

one base point, which can also be regarded as a (partial) generalization of a result of Severi [3] (proved by transcendental methods).

1. Let U and V be topological spaces, and let $G(y)$ be a function which assigns to each point y in V a subset $G(y)$ in U. We shall say that the system of subsets $G(y)$ defines a fibre system in U, if there exists an open covering $N=\{N\}$ of V such that for each set N there exists a continuous function $\phi_N(x,y)$, defined for all points $x \subset G(N)$, $y \subset N$ in the product space $U \times V$ with values in U, with the following properties:

$$\phi_N(x,y) \subset G(y), \qquad (x \subset G(N), y \subset N),$$

$$\phi_N(x,y) = x \qquad (x \subset G(y), y \subset N).$$

The space V is called the base space of the fibre system, and the open sets N and the corresponding functions $\phi_N(x,y)$ are called the slicing neighborhoods and the slicing functions respectively. Let $f(z)$ and $g(z)$ be continuous mappings of a topological space Z into U and V respectively, such that for each point $z \subset Z$ we have $f(z) \subset G(g(z))$, and let $g(z,t)$, $0 \leq t \leq 1$, be a homotopy of the mapping $g(z)$ in V. Then a homotopy $f(z,t)$, $0 \leq t \leq 1$, of $f(z)$ in U is said to cover the homotopy $g(z,t)$, or a covering homotopy of $g(z,t)$, if we have $f(z,t) \subset G(g(z,t))$ for all z, t. The covering homotopy theorem (in the weak form) asserts that if V is a normal Hausdorff space and if Z is compact, then there exists always a covering homotopy; furthermore, if $g(z,t)$ leaves a point $z_0 \subset Z$ fixed, we can assume that $f(z,t)$ also leaves z_0 fixed. That this covering homotopy theorem is true for any fibre system in U can be seen as follows. We observe first that in case the $G(y)$ is the inverse function $\pi^{-1}(y)$ of a continuous mapping $\pi(x)$ of U onto V, then we have a (generalized) fibre space as defined by S. T. Hu [4]; and, as has been observed by Hu, the covering homotopy theorem is true for such a fibre space. The general case can be reduced to this special case by considering the graph W of the function $G(y)$, i. e. the set of all points $w = x \times y$ in $U \times V$ satisfying the condition $x \subset G(y)$. Let $\tau(w)$ and $\pi(w)$ be the

[3] F. Severi, "Intorno al teorema d'Abel sulle superficie algebriche ed alla riduzione a forma normale degl'integrali di Picard," *Rendiconti del Circolo Matematico di Palermo*, vol. 21 (1906), p. 261, Teorema I.

[4] Sze-Tsen Hu, "On generalizing the notion of fibre spaces to include the fibre bundles," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 756-762. It is convenient to use this generalized notion of a fibre space, though the particular fibre systems used in the present paper can all be "derived" from fibre bundles which are also at the same time fibre spaces in the sense of Hurewicz-Steenrod.

mappings of W into U and V induced by the projections of $U \times V$ into U and V respectively. It is clear that $\pi(w)$ is a mapping of W onto V, and that we have $\tau(\pi^{-1}(y)) = G(y)$ for each $y \subset V$. The space W can then be made into a fibre space with respect to the mapping $\pi(w)$ if we define for each N the slicing function by the formula:

$$\phi_N(w, y) = \phi_N(\tau(w), y) \times y \qquad (w \subset \pi^{-1}(N), y \subset N).$$

If we set $h(z) = f(z) \times g(z)$, then $h(z)$ is a continuous mapping of Z into W such that $h(z) \subset \pi^{-1}(g(z))$. Since the covering homotopy theorem holds for the fibre mapping $\pi(w)$, there exists a homotopy $h(z, t)$ of $h(z)$ in W such that $h(z, t) \subset \pi^{-1}(g(z, t))$. Then we have $\tau h(z, t) \subset \tau\pi^{-1}(g(z, t)) = G(g(z, t))$, so that $\tau h(z, t)$ is a homotopy of $f(z)$ in U which covers the homotopy $g(z, t)$ in V.

[*Note added in proof* (June 6, 1952). Professor Beno Eckmann has recently called my attention to the fact that essentially the same concept as that of a fibre system has been introduced by him under the name "retrahierbare Ueberdeckung" in his paper "Zur Homotopietheorie gefaserter Raeume," *Commentarii Mathematici Helvetici*, vol. 14 (1941), pp. 141-192. In the first part of that paper the covering homotopy theorem was proved for a "retrahierbare Ueberdeckung", and an application was made of this theorem to the covering of a sphere by its great spheres.]

Let U be a non-singular algebraic variety of dimension r, and let Φ be an irreducible algebraic correspondence of dimension t between U and an algebraic variety V of dimension s. Then for a generic point η in V the set $\Phi^{-1}(\eta)$ is an irreducible algebraic variety of dimension $d = t - s$ in U, and we can consider $\Phi^{-1}(\eta)$ as a generic element of an algebraic system of d-cycles in U. For any point y in V the variety $\Phi^{-1}(y)$ is the carrier of the set of all d-cycles which are specializations of the d-cycle $\Phi^{-1}(\eta)$ over the specialization $\eta \rightarrow y$. A point y in V is said to be semiregular with respect to the correspondence Φ (or rather the inverse correspondence Φ^{-1}), if there is a uniquely determined specialization cycle of $\Phi^{-1}(\eta)$ over the specialization $\eta \rightarrow y$, and if this specialization cycle has no multiple components. It is easily seen that a point y in V is semi-regular with respect to Φ if and only if the variety $\Phi^{-1}(y)$ has the dimension d and the same degree (in the ambient projective space) as the variety $\Phi^{-1}(\eta)$, so that we can consider $\Phi^{-1}(y)$ itself as the specialization cycle of $\Phi^{-1}(\eta)$ over the specialization $\eta \rightarrow y$.

Since U is a differentiable manifold (of class C^∞), we can introduce in U a Riemannian metric. In fact, let $M = \{M_i\}$ be a locally finite system of

coordinate neighborhood covering U, and let the set of differentiable functions $\{e_i(x)\}$ be a partition of unity subordinate to this covering; then, if we denote by ds_i^2 the Euclidean metric of the coordinate neighborhood M_i, the differential form $ds^2 = \sum_i e_i(x) ds_i^2$ defines a Riemannian metric on U. We observe that by means of a suitable choice of the covering neighborhood M_i and the partition functions $e_i(x)$ we can make the Riemannian metric ds^2 in a sufficiently small neighborhood of any given point equal to the Euclidean metric with respect to any given coordinate system around this point. This fact will be convenient for us later. Let y' be a semi-regular point in V, and let R be a compact subset in $\Phi^{-1}(y')$ consisting of only simple points in $\Phi^{-1}(y')$. If $p(x)$ is any differentiable function which assigns to each point x in R a $(2r-2d)$-dimensional direction element which is transversal to the tangent space of $\Phi^{-1}(y')$ at the point x, then there exists a differentiable system of $(2r-2d)$-dimensional geodesic surfaces (or geodesic $(2r-2d)$-surfaces) $P(x)$ $(x \subset R)$, such that for each point x in R the surface $P(x)$ has the tangential direction $p(x)$ at x. If N is a sufficiently small neighborhood of y' in V, then for every point y in N and every point x in R, the intersection $\Phi^{-1}(y) \cap P(x)$ consists of exactly one point which is simple in $\Phi^{-1}(y)$, and the mapping $x \to \Phi^{-1}(y) \cap P(x)$ is a homeomorphism of R onto a compact subset $R(y)$ in $\Phi^{-1}(y)$. Thus each point x in $\sum_{y \subset N} R(y)$ is contained in exactly one geodesic $(2r-2d)$-surface of the system, which we shall also denote by $P(x)$; the function $\phi_N(x, y) = P(x) \cap R(y)$ $(x \subset \sum_{y \subset N} R(y), y \subset N)$ is then a slicing function for the system $R(y)$ $(y \subset N)$.

In case the variety $\Phi^{-1}(y')$ is non-singular, we can set $R = \Phi^{-1}(y')$ and hence $R(y) = \Phi^{-1}(y)$ for all y in N, so that the system of varieties $\Phi^{-1}(y)$ $(y \subset N)$ defines a fibre system in U. Now, if the generic variety $\Phi^{-1}(\eta)$ *is* non-singular, then there exists a proper subvariety H in V such that every point y in $V-H$ is semi-regular with respect to Φ and the variety $\Phi^{-1}(y)$ is non-singular. It follows then that the system of varieties $\Phi^{-1}(y)$ $(y \subset V-H)$ defines a fibre system in U. We shall now show that in case $d=1$ this assertion is also true (with a suitable definition of H) even if $\Phi^{-1}(\eta)$ has singular points. We shall say that a point y' in V is regular with respect to Φ, if y' is semi-regular and if the variety $\Phi^{-1}(y')$ has the same singularities as the generic variety $\Phi^{-1}(\eta)$, i. e. each singular branch of the curve $\Phi^{-1}(y')$ is the specialization of a singular branch of the same order of the curve $\Phi^{-1}(\eta)$. It is easily seen that the set of all points in V which are not regular with respect to Φ is a proper subvariety in V, which we shall also denote by H. Let y' be a point in $V-H$, and let x' be a singular point in $\Phi^{-1}(y')$; then

there is a coordinate neighborhood M of x' in U and a suitably chosen system of coordinates $u_1, \cdots u_r$ in M with origin at x', such that for every point y in a sufficiently small neighborhood N of y' in V, the curve $\Phi^{-1}(y) \cap M$ is a regular analytic covering space of a fixed degree g over a neighborhhood M_1 of the origin in the complex u_1-plane with a unique branch point over the origin $u_1 = 0$. Let M'_1 be a circular region $|u_1| < \epsilon$ $(\epsilon > 0)$ such that its closure $\bar{M}'_1$ is contained in M_1, and let $P(\alpha)$ $(\alpha \subset M_1)$ be the "hyperplane" in M defined by the equation $u_1 = \alpha$. Then, for each $u_1 \neq 0$ in $\bar{M}'_1$, there exist g disjoint circular domains $P_i(u_1)$, $i = 1, \cdots, g$, in $P(u_1)$, such that for each y in N and for each $i = 1, \cdots, g$, the intersection $\Phi^{-1}(y) \cap P_i(u_1)$ consists of exactly one point, while for $u_1 = 0$, the intersection $\Phi^{-1}(y) \cap P(0)$ itself consists of exactly one point, namely the branch point of $\Phi^{-1}(y)$ over $u_1 = 0$. For each y in N, we set $L(y) = \sum_{u_1 \subset M'_1} \Phi^{-1}(y) \cap P(u_1)$ and $\bar{L}(y) = \sum_{u_1 \subset \bar{M}'_1} \Phi^{-1}(y) \cap P(u_1)$ so that $\bar{L}(y)$ is the closure of the domain $L(y)$ in $\Phi^{-1}(y)$. We can then define a slicing function for the system $\bar{L}(y)$ $(y \subset N)$ by setting $\phi_N(x, y) = \Phi^{-1}(y) \cap P_i(u_1)$, $(x \subset \sum_{y \subset N} \bar{L}(y), y \subset N)$, where $P_i(u_1)$ is the one of the g domains in $P(u_1)$ which contains the point x. Furthermore, if we choose our Riemannian metric in U in such a way that it coincides with the Euclidean metric in the coordinate neighborhood M, then each $P_i(u_1)$ is a geodesic $(2r - 2)$-surface, and for each point $x \neq x'$ in $\bar{L}(y')$ the $P_i(u_1)$ passing through it has a tangential direction element $p(x)$ which is transversal to the tangential space of $\Phi^{-1}(y')$ at x. In particular this function $p(x)$ is defined on the boundary $\bar{L}(y') - L(y')$ curve of the domain $L(y')$, and it is also differentiable. Let now $x^{(i)}$, $i = 1, \cdots, a$, be the singular points of $\Phi^{-1}(y')$, and let $M^{(i)}$, $i = 1, \cdots, a$, be suitably chosen (disjoint) coordinate neighborhoods of the points $x^{(i)}$, $i = 1, \cdots, a$, respectively in U, and let the Riemannian metric in U be so chosen that it coincides with the Euclidean metric in each $M^{(i)}$. If we choose the neighborhood N of y' sufficiently small, then we can define a fibre system $\bar{L}^{(i)}(y)$ $(y \subset N)$ and a slicing function $\phi^{(i)}{}_N(x, y)$ $(x \subset \sum_{y \subset N} \bar{L}^{(i)}(y), y \subset N)$ in each $M^{(i)}$, and we have also the function $p^{(i)}(x)$ defined in $\bar{L}^{(i)}(y') - L^{(i)}(y')$. If we set $Q(y) = \sum_{i=1}^{a} L^{(i)}(y)$ and $\bar{Q}(y) = \sum_{i=1}^{a} \bar{L}^{(i)}(y)$, then we can consider all the functions $\phi^{(i)}{}_N(x, y)$ together as a single slicing function $\phi_N(x, y)$ $(x \subset \sum_{y \subset N} \bar{Q}(y), y \subset N)$ for the fibre system $\bar{Q}(y)$ $(y \subset N)$, and also all the functions $p^{(i)}(x)$ together as a single function $p(x)$ defined on the boundary $\bar{Q}(y') - Q(y')$ of the domain $Q(y')$. For each $y \subset N$, let $R(y) = \Phi^{-1}(y) - Q(y)$; it is clear that $\bar{Q}(y)$

and $R(y)$ are two complementary closed domains in $\Phi^{-1}(y)$ with the curve $\bar{Q}(y) - Q(y)$ as their common boundary. In particular the boundary of $R(y')$ is $\bar{Q}(y') - Q(y')$, and the function $p(x)$ is defined and differentiable on the boundary $\bar{Q}(y') - Q(y')$. If we assign to each point x in $R(y')$ the set of all $(2r-2)$-dimensional direction elements at x which are transversal to the tangent space of $\Phi^{-1}(y')$ at x, then we obtain a differentiable fibre bundle over the space $R(y')$, in which the fibre is topologically a $(4r-4)$-cell. It follows then that the differentiable function $p(x)$, which is defined on the boundary of $R(y')$, can be extended to a differentiable function $p'(x)$ in the entire space $R(y')$. If we denote by $P'(x)$ $(x \subset R(y'))$ the system of geodesic $(2r-2)$-surfaces corresponding to the function $p'(x)$, then for each y in N, provided N is taken sufficiently small, the mapping $x \rightarrow \Phi^{-1}(y) \cap P'(x)$ is a homeomorphism of $R(y')$ onto a closed domain in $\Phi^{-1}(y)$ whose boundary is $\bar{Q}(y) - Q(y)$ and which approaches $R(y')$ as y approaches y'; hence this domain must be $R(y)$. If we denote by $P'(x)$, for any point x in $\sum_{y \subset N} R(y)$, the one geodesic $(2r-2)$-surface of this system which contains the point x, then the function

$$\phi'_N(x,y) = R(y) \cap P'(x) = \Phi^{-1}(y) \cap P'(x) \quad (x \subset \sum_{y \subset N} R(y), y \subset N)$$

is a slicing function of the system $R(y)$ $(y \subset N)$. Since, for

$$x \subset (\sum_{y \subset N} R(y)) \cap (\sum_{y \subset N} \bar{Q}(y)),\ y \subset N,$$

we have

$$\phi_N(x,y) = \Phi^{-1}(y) \cap P_i(u_1) = \Phi^{-1}(y) \cap P'(x) = \phi'_N(x,y),$$

the two slicing function $\phi_N(x,y)$ and $\phi'_N(x,y)$ are concordant whenever both are defined and hence can be considered together as one slicing function for the system of curves $\Phi^{-1}(y)$ $(y \subset N)$. Thus we have shown that the system of curves $\Phi^{-1}(y)$ $(y \subset V - H)$ is a fibre system.

2. THEOREM 1. *Let U be a non-singular algebraic variety of dimension r, and let a subvariety G of dimension d be a member of an irreducible algebraic system $G(y)$ $(y \subset V)$, with U as its carrier variety, which has at least one base point, and whose generic member is irreducible. If $f(z)$ is a continuous mapping of the unit interval I into U, with $f(0) = f(1) = x^{(0)}$ in G, then a finite power of this mapping is homotopic* rel. $z = 0, 1$, *to a continuous mapping of I into G. If the system $G(y)$ is involutional, then $f(z)$ itself is homotopic* rel. $z = 0, 1$, *to a continuous mapping of I into G.*

Remark. The m-th power of $f(z)$ is the mapping $\bar{f}(z)$ of I defined by setting $\bar{f}(z) = f(mz - i)$ for $i/m \leq z \leq (i+1)/m$, $i = 0, 1, \cdots, m-1$. The system $G(y)$ is said to be involutional, if it is induced by a rational transformation of U onto V.

Proof. Let the algebraic system be defined by an irreducible correspondence Φ between U and the variety V, and let $y^{(0)}$ be the point in V such that $\Phi^{-1}(y^{(0)}) = G(y^{(0)}) = G$. It is sufficient to prove our theorem for the case where $y^{(0)}$ is any point in an everywhere dense subset in V; this follows from the fact that for any point y in V the variety $\Phi^{-1}(y)$ is a neighborhood deformation retract in U. We begin with the special case where the system $G(y)$ is the linear system cut out on U by the system of linear subspaces of dimension $n - r + d$ $(d \geq 1)$ in the ambient space S_n, which all pass through a sufficiently general $S_{n-r+d-1}$. Since in this case $\Phi^{-1}(\eta)$ is non-singular for a generic point η in V, there exists a subvariety H in V such that each point y in $V - H$ is semi-regular and $\Phi^{-1}(y)$ is non-singular. We can assume without any loss of generality that $y^{(0)}$ is a point in $V - H$; furthermore, we can also assume that $x^{(0)}$ is a point in G outside of $\Phi^{-1}(H)$. Since $\Phi^{-1}(H)$ is a proper subvariety in U and hence topologically a subcomplex of dimension $\leq 2r - 2$ in the $2r$-dimensional topological manifold U, we can assume that, after a suitable homotopy (rel. $z = 0, 1$) if necessary, $f(z)$ is a mapping of I into $U - \Phi^{-1}(H)$. Then the mapping $g(z) = \Phi(f(z))$ of I into $V - H$ is well defined, and we have evidently $g(0) = g(1) = y^{(0)}$ and $f(z) \subset G(g(z))$ for all z. Let $x^{(1)}$ be any point in $G \cap S_{n-r+d-1}$, and let $h(z)$ be any continuous mapping of I into G such that $h(0) = x^{(0)}$ and $h(1) = x^{(1)}$. We set

$$f'(z) = \begin{cases} f(4z) & (0 \leq z \leq \frac{1}{4}), \\ h(4z - 1) & (\frac{1}{4} \leq z \leq \frac{1}{2}), \\ x^{(1)} & (\frac{1}{2} \leq z \leq \frac{3}{4}), \\ h(4 - 4z) & (\frac{3}{4} \leq z \leq 1), \end{cases}$$

and

$$g'(z) = \begin{cases} g(4z) & (0 \leq z \leq \frac{1}{4}), \\ y^{(0)} & (\frac{1}{4} \leq z \leq \frac{1}{2}), \\ g(3 - 4z) & (\frac{1}{2} \leq z \leq \frac{3}{4}), \\ y^{(0)} & (\frac{3}{4} \leq z \leq 1); \end{cases}$$

it is clear that $f'(z) \simeq f(z)$ rel. $z = 0, 1$, and $g'(z) \simeq 0$ rel. $z = 0, 1$. Since $f'(z) \subset G(g'(z))$ for all z, it follows from the covering homotopy theorem that there is a homotopy of $f'(z)$ rel. $z = 0, 1$, which deforms $f'(z)$ into a mapping of I into G; hence $f(z)$ is also homotopic rel. $z = 0, 1$, to a mapping of I into G.

Turning to the general case, we observe first that we can assume without any loss of generality the following: (1) For a generic point ξ in U the variety $\Phi(\xi)$ consists of a finite number m of points; for otherwise we can replace V by its intersection with a suitably chosen linear subspace (passing through $y^{(0)}$) in its ambient space. It is clear that we have $m=1$ in case $G(y)$ is an involutional system. (2) For a generic point η in V the variety $\Phi^{-1}(\eta)$ is a curve, i. e. $d=1$; for otherwise we can replace U by its intersection with a suitably chosen S_{n-d+1} (passing through a base point of the system $G(y)$) in S_n, and we have just shown above that any continuous mapping $f(z)$ of I into U, with $f(0)=f(1)$ in $U \cap S_{n-d+1}$, is homotopic rel. $z=0,1$, to a continuous mapping of I into $U \cap S_{n-d+1}$. Now, let H be the subvariety in V containing all points which are not regular with respect to Φ, so that the system of curves $G(y)$ $(y \subset V-H)$ is a fibre system; let T be the subvariety in U such that for every point x in $U-T$ the set $\Phi(x)$ consists of m distinct points outside of H. Without any loss of generality we can assume that $y^{(0)}$ is a point in $V-H$ such that $G=\Phi^{-1}(y^{(0)})$ is not entirely in T, and that $x^{(0)}$ is a point in $G-T$; then $\Phi(x^{(0)})$ consists of m points, one of which is the point $y^{(0)}$. Since $T+\Phi^{-1}(H)$ is a proper subvariety in U and hence topologically a subcomplex of dimension $\leq 2r-2$ in the $2r$-dimensional manifold U, we can assume that, after a suitable homotopy (rel. $z=0,1$) if necessary, $f(z)$ is a mapping of I into $U-T-\Phi^{-1}(H)$. Then the image $\Phi(f(z))$ consists of m distinct mappings of I into $V-H$, one (and only one) of which will be a mapping $g(z)$ such that $g(0)=y^{(0)}$. The point $g(1)$ is one of the m points in the set $\Phi(x^{(0)})$, though not necessarily the point $y^{(0)}$. It is easily seen that if $\bar{f}(z)$ is the m-th power of $f(z)$, then the image $\Phi(\bar{f}(z))$ of $\bar{f}(z)$ will consist of m distinct mappings of I into $V-H$, one of which will be a mapping $\bar{g}(z)$ such that $\bar{g}(0)=\bar{g}(1)=y^{(0)}$. Let $x^{(1)}$ be a base point of the system $G(y)$, and let $h(z)$ be a continuous mapping of I into G such that $h(0)=x^{(0)}$ and $h(1)=x^{(1)}$. If we now define $f'(z)$ and $g'(z)$ again as before, replacing $f(z)$ and $g(z)$ by $\bar{f}(z)$ and $\bar{g}(z)$ respectively, we can repeat exactly the same argument and conclude that $\bar{f}(z)$ is homotopic rel. $z=0,1$, to a mapping of I into G. This concludes the proof of Theorem 1.

In the following we shall denote by $F(U)$ the fundamental group of a topological space U, considered as a group of mapping classes with some one fixed reference point. If W and X are two subsets in U, then the identity mapping of $W \cap X$ into W will induce a homomorphism of $F(W \cap X)$ into $F(W)$, the reference point in both groups being one and the same point in $W \cap X$; we shall then denote by $F(W,X)$ the subgroup of $F(W)$ which is the image of $F(W \cap X)$ under this homomorphism.

THEOREM 2. *Let Φ be a rational transformation of a non-singular algebraic variety U of dimension r onto a non-singular algebraic variety V of dimension s, with the properties:* (1) *For a generic point η in V the variety $\Phi^{-1}(\eta)$ is irreducible, and* (2) *the set H of all points which are not semi-regular with respect to Φ is a subvariety of dimension $\leq s-2$ in V. Then there is a homomorphism of $F(U)$ onto $F(V)$, and the kernel of this homomorphism is the subgroup $F(U, \Phi^{-1}(y^{(0)}))$, where $y^{(0)}$ is any sufficiently general point in V.*

Proof. Let T be the fundamental variety of Φ in U; it is well known that T has a dimension $\leq r-2$. It is clear that the variety $\Phi^{-1}(\eta)$ has the dimension $d=r-s$; let c be the degree of $\Phi^{-1}(\eta)$ considered as a variety in the ambient space S_n of U. Let $y^{(0)}$ be a point in V such that $\Phi^{-1}(y^{(0)})$ is not contained in T, and let $x^{(0)}$ be a point in $\Phi^{-1}(y^{(0)})-T$. We shall take $x^{(0)}$ as the reference point of the groups $F(U)$, $F(U-T)$, $F(\Phi^{-1}(y^{(0)}))$, $F(U, \Phi^{-1}(y^{(0)}))$, and $F(U-T, \Phi^{-1}(y^{(0)}))$, and take $y^{(0)}$ as the reference point of the group $F(V)$. The identity mapping of $U-T$ into U induces a homomorphism θ of $F(U-T)$ into $F(U)$; since T is topologically a subcomplex of dimension $\leq 2r-4$ in the $2r$-dimensional manifold U, it is easily seen that θ is an isomorphism of $F(U-T)$ onto $F(U)$, and that the image of $F(U-T, \Phi^{-1}(y^{(0)}))$ under θ is precisely $F(U, \Phi^{-1}(y^{(0)}))$. Since Φ is a continuous mapping of $U-T$ into V, there is an induced homomorphism ϕ of $F(U-T)$ into $F(V)$. We set $\Delta=\phi\theta^{-1}$, so that Δ is a homomorphism of $F(U)$ into $F(V)$. We shall prove our theorem by showing that (provided the point $y^{(0)}$ is sufficiently general) Δ is a homomorphism of $F(U)$ onto $F(V)$ and its kernel is $F(U, \Phi^{-1}(y^{(0)}))$.

Let L be a linear subspace of dimension $n-r+s$ in the ambient space S_n of U, such that $U \cap L$ is a non-singular variety of dimension s. The rational transformation Φ will then induce a rational transformation $\bar{\Phi}$ of $U \cap L$ onto V, and we have $\bar{\Phi}^{-1}(y) \subset \Phi^{-1}(y) \cap L$ for every point y in V. Since, for a general point η in V, the set $\bar{\Phi}^{-1}(\eta)=\Phi^{-1}(\eta) \cap L$ consists of c distinct points in $U-T$, there exists a proper subvariety K in V such that for every point y in $V-K$ the set $\bar{\Phi}^{-1}(y)$ consists of c distinct points in $U-T$ (which are then necessarily simple points in the variety $\Phi^{-1}(y)$). We shall also assume that $y^{(0)}$ is a point in $V-K$ and $x^{(0)}$ is a point in $\bar{\Phi}^{-1}(y^{(0)})$.

Let $g(z)$ be a continuous mapping of the unit interval I into V such that $g(0)=g(1)=y^{(0)}$. Since K is topologically a subcomplex of dimension $\leq 2s-2$ in the topological manifold V of dimension $2s$, we can assume, after

a homotopy (rel. $z=0,1$) if necessary, that $g(z)$ is a mapping of I into $V-K$. Since $\bar{\Phi}^{-1}(V-K)$ is a c-fold regular covering space of $V-K$, the inverse image $\bar{\Phi}^{-1}(g(z))$ consists of c distinct mappings of I into $U-T$, one of which will be a mapping $f(z)$ such that $f(0)=x^{(0)}$. Let $x^{(1)}=f(1)$, and let $f_1(z)$ be any continuous mapping of I into $\Phi^{-1}(y^{(0)})-T$ such that $f_1(0)=x^{(1)}$ and $f_1(1)=x^{(0)}$. If we set

$$f'(z)=\begin{cases} f(2z) & (0\leq z\leq \tfrac{1}{2}),\\ f_1(2z-1) & (\tfrac{1}{2}\leq z\leq 1),\end{cases}$$

then we have evidently a continuous mapping of I into $U-T$ such that $\Phi(f'(z))\simeq g(z)$ rel. $z=0,1$. This shows that Δ is a homomorphism of $F(U)$ onto $F(V)$.

Now let $f(z)$ be a continuous mapping of I into U such that $f(0)=f(1)=x^{(0)}$. According to Theorem 1, $f(z)$ is homotopic rel. $z=0,1$ to a mapping of I into $U\cap L$, and hence also homotopic rel. $z=0,1$ to a mapping of I into $\bar{\Phi}^{-1}(V-K)$. We can therefore assume that $f(z)$ is already a mapping of I into $\bar{\Phi}^{-1}(V-K)$; then the mapping $g(z)=\Phi(f(z))$ is defined, and we have $g(0)=g(1)=y^{(0)}$. In order to prove that the kernel of Δ is $F(U,\Phi^{-1}(y^{(0)}))$, we have only to show that if the mapping $g(z)$ is homotopic to zero rel. $z=0,1$ in V, then $f(z)$ is homotopic rel. $z=0,1$ to a mapping of I into $\Phi^{-1}(y^{(0)})$. Without loss of generality we can restrict ourselves to the case $d=1$; for otherwise we can replace U by its intersection with a suitably chosen linear subspace of dimension $n-r+s+1$ which contains the space L. It is then easily seen that we can assume the subvariety K so chosen that every point in $V-K$ is regular with respect to Φ; then the system of curves $\Phi^{-1}(y)$ $(y\subset V-K)$ is a fibre system. Furthermore, there exists a subvariety H' of dimension $\leq s-2$ in V, which contains H and is contained in K, such that for every point y in $V-H'$ the set $\Phi^{-1}(y)$ consists of at most c points, each of which is either a simple point in $\Phi^{-1}(y)$ or a singular point of a "generic" nature.

Since K is topologically a subcomplex of dimension $\leq 2s-2$, and H' is topologically a subcomplex of dimension $\leq 2s-4$ in the $2s$-dimensional manifold V, there exists in $V-H'$ a homotopy $g(z,t)$ of the mapping $g(z)$ rel. $z=0,1$ with the following properties: (1) $g(z,t)\subset V-K$ for all z and t except a finite number m of points $g((2i-1)/2m,1)$, $i=1,\cdots,m$, which all belongs to $K-H'$; (2) $g(i/m,1)=y^{(0)}$ for $i=1,\cdots,m$, and $g(z,1)=g((2i-1-mz)/m,1)$ for $(i-1)/m\leq z\leq i/m$. Since $\bar{\Phi}^{-1}(V-H')$ is an analytic covering space (with branch points) of $V-H'$, and since moreover $\bar{\Phi}^{-1}(V-K)$ is a regular covering space of $V-K$, there corre-

sponds to the homotopy $g(z, t)$ a uniquely determined homotopy $f(z, t)$ rel. $z = 0, 1$ in $\bar{\Phi}^{-1}(V - H')$, with $f(z, 0) = f(z)$ and $f(z, t) \subset \bar{\Phi}^{-1}(g(z, t))$. Therefore, to prove our assertion, it is sufficient to consider the following problem. Let $g(z)$ be a continuous mapping of I into $V - H'$ with the properties: $g(0) = g(1) = y^{(0)}$; $g(z) = g(1 - z)$ for all z; $g(z) \subset V - K$ for all $z \neq \frac{1}{2}$, and $g(\frac{1}{2})$ belongs to $K - H'$. Let $f(z)$ be a continuous mapping of I into $\bar{\Phi}^{-1}(V - H')$ such that $f(z) \subset \bar{\Phi}^{-1}(g(z))$ for all z; then we have to show that $f(z)$ is homotopic rel. $z = 0, 1$ in U to a mapping of I into $\Phi^{-1}(y^{(0)})$.

Let $x^{(1)} = f(\frac{1}{2})$ and $y^{(1)} = g(\frac{1}{2})$, and let R be a compact neighborhood of $x^{(1)}$ in $\Phi^{-1}(y^{(1)})$ which contains no singular points of $\Phi^{-1}(y^{(1)})$ except possibly the point $x^{(1)}$ itself. Then, as we have shown in the preceding section, there exists a neighborhood N of $y^{(1)}$, and to every point y in N there corresponds a compact subset $R(y)$ in $\Phi^{-1}(y)$ (with $R(y^{(1)}) = R$), such that the system $R(y)$ $(y \subset N)$ is a fibre system; furthermore, there is a neighborhood M of $x^{(1)}$ in U such that $M \cap \Phi^{-1}(y) \subset R(y)$ for all y in N. Let I_e $(e > 0)$ denote the interval $|z - \frac{1}{2}| \leq e$; and let e be taken so small that $f(z) \subset M$ and $g(z) \subset N$ for all z in I_e. Since the mapping $g(z)$ of I_e into N is evidently homotopic rel. $z = \frac{1}{2} - e, \frac{1}{2} + e$ in N to a mapping of I_e into the point $y^{(2)} = g(\frac{1}{2} - e) = g(\frac{1}{2} + e)$, it follows from the covering homotopy theorem that the mapping $f(z)$ of I_e into $M \cap \Phi^{-1}(N)$ is also homotopic rel. $z = \frac{1}{2} - e, \frac{1}{2} + e$ to a mapping $h(z)$ of I_e into $\Phi^{-1}(y^{(2)})$. If we set

$$f'(z) = \begin{cases} f(z) & (|z - \frac{1}{2}| \geq e), \\ h(z) & (|z - \frac{1}{2}| \leq e), \end{cases}$$

$$g'(z) = \begin{cases} g(z) & (|z - \frac{1}{2}| \geq e), \\ y^{(2)} & (|z - \frac{1}{2}| \leq e), \end{cases}$$

then we have evidently $f(z) \simeq f'(z)$ rel. $z = 0, 1$ in U and $g'(z) \simeq 0$ rel. $z = 0, 1$ in $V - K$. Since $f'(z) \subset \Phi^{-1}(g'(z))$ for all z, it follows then from the covering homotopy theorem that $f'(z)$ is homotopic rel. $z = 0, 1$ to a mapping of I into $\Phi^{-1}(y^{(0)})$, and hence $f(z)$ is also homotopic rel. $z = 0, 1$ to a mapping of I into $\Phi^{-1}(y^{(0)})$. This concludes the proof of Theorem 2.

THE JOHNS HOPKINS UNIVERSITY.

INDUCED REPRESENTATIONS.*

By F. I. Mautner.

1. Introduction; statement of Theorem 1. Let G be a locally compact topological group and K an arbitrary (but fixed) compact subgroup of G. With every continuous unitary representation u of K in a Hilbert (or finite dimensional) space $\mathfrak{h}$ over the complex numbers we can associate a continuous unitary representation U of G, the so called *induced representation*, as follows:

Consider all those Haar-measurable functions $X(g)$ defined on G with values in $\mathfrak{h}$ for which [1]

$$\int_G \|X(g)\|^2 dg < \infty,$$

where $\|X(g)\|$ denotes, for fixed g, the norm of $X(g)$ as an element of the Hilbert space $\mathfrak{h}$. We restrict ourselves to only those functions $X(g)$ which satisfy also

$$X(gk) = u(k^{-1})X(g) \quad \text{for all } k \varepsilon K. \tag{1.1}$$

Clearly all such functions form a linear space over the complex numbers from which we obtain a Hilbert (or finite dimensional) space $\mathfrak{H}$ if we identify functions $X(g)$ which differ on sets of Haar-measure zero, and define an inner product (X, Y) in $\mathfrak{H}$ by

$$(X, Y) = \int_G (X(g), Y(g))dg, \tag{1.2}$$

where $(X(g), Y(g))$ denotes for, fixed $g \varepsilon G$, the inner product in the space $\mathfrak{h}$. Now define for every $\gamma \varepsilon G$ a linear transformation $U(\gamma)$ by

$$(U(\gamma)X)(g) = X(\gamma^{-1}g). \tag{1.3}$$

* Received April 19, 1951.

[1] Introduce in $\mathfrak{h}$ an arbitrary complete orthonormal system, and denote by $X_j(g)$ the expansion coefficients of $X(g)$ with respect to it (for each fixed $g \varepsilon G$). We shall say that the vector-valued function $X(g)$ is Haar-measurable if each of the complex valued functions $X_j(g)$ is Haar-measurable. It is clear that this definition is independent of the particular complete orthonormal system in $\mathfrak{h}$, and that it follows that the inner product $(X(g), Y(g))$ of any two Haar-measurable vector valued functions $X(g)$ and $Y(g)$ is Haar-measurable.

Clearly $U(\gamma)$ is a unitary operator of the space $\mathfrak{H}$ onto itself, and the mapping $U: \gamma \to U(\gamma)$ defines a continuous unitary representation of G in the space $\mathfrak{H}$. We call U the *induced representation generated by u,* and write

$$(1.4) \qquad U = \operatorname{ind} u \quad \text{or} \quad U = \operatorname*{ind}_{K\uparrow G} u.$$

For finite groups this is the same as the classical definition of "induced representation." For an arbitrary locally compact group and a closed subgroup a definition has recently been given by Mackey [5]. It is easy to see that Mackey's definition reduces to the above in the case when the subgroup K is compact, which we shall assume throughout.

Suppose for the moment that G is compact too. Then the representation U of G can be decomposed into a discrete (i. e. ordinary) direct sum $\sum_{\oplus} \mu_j M_j$ of irreducible finite dimensional pairwise inequivalent representations M_j of G, each occurring with multiplicity μ_j. The classical Frobenius reciprocity theorem asserts in this case that

$$(1.5) \qquad \text{multiplicity of } u \text{ in } M_j(K) = \mu_j,$$

where $M_j(K)$ denotes the restriction of the representation M_j to the subgroup K.

The problem arises whether this theorem can be generalized to the case where G is no longer compact. It is known that U can still be decomposed into irreducible unitary representations M_j in the sense of generalized direct sums (= direct integrals). Therefore equation (1. 5) can still be formulated. However [2] there exist infinite discrete groups for which (1. 5) is false even when one takes for K the trivial subgroup = (1).

But one can replace (1. 5) by the following formulation, which is equivalent to (1. 5) in the classical case (G compact). Denote by U_j the repetition μ_j times of M_j:

$$U_j = M_j \oplus M_j \oplus M_j \oplus \cdots \oplus M_j.$$

Then we have $U = \sum_{\oplus} U_j$, where U_i and U_j are inequivalent for $i \neq j$. Denote by W_j the algebra generated by the operators $U_j(g)$ in the representation space $\mathfrak{H}_j$ of U_j, and by W'_j the commuting algebra of W_j in $\mathfrak{H}_j$. Then the classical theory of linear algebras tells us that

---

[2] G. W. Mackey has found an explicit decomposition of the regular representation of certain discrete groups for which every irreducible component representation occurs with multiplicity one, and is infinite dimensional (oral communication of an unpublished result).

$$\dim(W'_j) = \mu_j^2, \tag{1.6}$$

where $\dim(W'_j)$ = dimension of W'_j as a linear space over the complex numbers. Then we have

$$\text{multiplicity of } u \text{ in } U_j(K) = \dim(W'_j). \tag{1.7}$$

Moreover it is well known that if $E_1, E_2, \cdots$ denote the minimal self-adjoint idempotents in the center Z of W, then $W_j = E_jW$, and $W'_j = E_jW'$. This suggests another possibility of generalization, namely the use of von Neumann's central decomposition. And, in fact, form (1.7) of the Frobenius reciprocity theorem can be generalized as follows. Assume now that G is an arbitrary locally compact group whose Haar measure is both left and right invariant, and which satisfies the second aviom of countability. Then the above Hilbert space $\mathfrak{H}$ is separable, so we can apply Theorem VII of [15]. Indeed, let W be the weakly closed self-adjoint algebra of bounded linear operators generated by the operators $U(g)$ in $\mathfrak{H}$, W' the commuting algebra of W, and Z the center,, i. e. $Z = W \cap W'$. Then we obtain a direct integral decomposition

$$\mathfrak{H} = \int_{\oplus} \mathfrak{H}_t \tag{1.8}$$

under the operators $U(g)$ to which the center Z "belongs" in the sense af Theorems IV and VII of [15]. For each $g \in G$ we obtain an operator-valued function $U(g, t)$ of t which can be changed arbitrarily on t-sets of measure zero. It has been proved in Theorem 1.1 of [8] that one can find for each $g \in G$ one such operator-valued function $U(g, t)$, and for each t a continuous unitary repersentation $\mathfrak{V}_t$: $g \to \mathfrak{V}_t(g)$ of G in the space $\mathfrak{H}_t$ such that

$$U(g, t) = \mathfrak{V}_t(g) \quad \text{for } g \notin N_t, \tag{1.9}$$

where N_t is some subset of G of Haar measure zero, depending on t. Then we have

THEOREM 1. *Let G be a locally compact unimodular group, satisfying the second axiom of countability. Let K be an arbitrary compact subgroup of G and u: $k \to u(k)$ a continuous irreducible (unitary) representation of the subgroup K. Perform the central decomposition* (1.8) *of the space $\mathfrak{H}$ of the induced representation $U = \operatorname{ind} u$. Denote by $\mathfrak{V}_t(K)$ the restriction of the representation $\mathfrak{V}_t$ of G to the subgroup K, and by $[\mathfrak{V}_t(g)]'$ the algebra of all those bounded operators in $\mathfrak{H}_t$ which commute with $\mathfrak{V}_t(g)$ for every $g \in G$.*

ASSERTION.

(1.10) *Multiplicity of* u *in* $\mathfrak{V}_t(K) = \dim\{[\mathfrak{V}_t(g)]'\}$ *for almost every* t,

where "dim" *denotes the ordinary dimension of a linear space over the complex numbers.*

This theorem will be proved in **2** to **5**. Some of the lemmas proved in **2** to **5** may also be of interest independently of the proof of Theorem 1. There is some overlap between Theorem 1 above and the generalizations of the Frobenius reciprocity theorem recently obtained by Mackey [6]. The restrictions under which we prove Theorem 1 above are different from the assumption under which Mackey's results hold.

This suggests the possibility of a more general result which should contain all the known generalizations of the Frobenius reciprocity theorem. We shall not discuss this further generalization, which seems to present serious difficulties, in the present paper. However Theorem 1 in its present form has various applications. The simplest application is to the case where the commuting algebra W' of a given induced representation ind u is known to be commutative. For W' commutative implies $\dim\{[\mathfrak{V}_t(g)]'\} = 1$ for almost every t; hence by Theorem 1 the multiplicity of u in $\mathfrak{V}_t(K)$ equals 1 for almost every t. This result is the main step in the derivation of the Plancherel formula outlined in [11a]. The details of this derivation were included in the original version of this paper, but have been separated from the rest of the paper at the suggestion of the referee.

Theorem 1 can also be applied to the case where G is a semi-simple Lie Group, which for simplicity of statement we take to be its own adjoint group, and K a maximal compact subgroup of G. The results of Harish-Chandra together with Theorem 1 above imply in this case that the above commuting algebra W' is a finite module over its center for every irreducible continuous representation u of K. This implies that the methods and results outlined in [11a] apply to the representation space $\mathfrak{H}$ of ind u in this case for every u. Thus one obtains by very general considerations a rather special kind of Plancherel formula for each $\mathfrak{H}$, and hence also for the space $\mathfrak{L}_2(G)$ of all complex valued Haar-Lebesgue square integrable functions on G.

We shall use in this paper essentially the same terminology and notation as in [7] and [8]. In particular Hilbert spaces are always assumed to admit complex scalars and will be denoted by $\mathfrak{H}$, $\mathfrak{h}$ or $\mathfrak{H}_t$, and occasionally also by $\boldsymbol{H}$ or $\boldsymbol{H}_t$. Whenever we use the results of [15] in an essential manner the Hilbert spaces in question have to be assumed to be separable (they may be

finite dimensional). For the notions and properties of generalized direct sums $\mathfrak{H} = \int_{\oplus} \mathfrak{H}_t$ we refer the reader to [15]. If we are given such a generalized direct sum (= direct integral), then there corresponds to every $x \varepsilon \mathfrak{H}$ a vector valued function $x(t)$ such that the value of $x(t)$ is for a given t an element of $\mathfrak{H}_t$. We call $x(t)$ the "*component*" of x in the space $\mathfrak{H}_t$ and say also that "*x decomposes into $x(t)$*." To define the direct integrals the points t must form a measure space. Our assertions about a given direct integral will always be about "almost all t" or "almost all the spaces $\mathfrak{H}_t$" etc. by which will be meant "all t (or $\mathfrak{H}_t$ etc.) except for a set of elements t whose measure (with which the given direct integral is formed) is zero." I. e. the measure referred to in connection with a given direct integral will always be the particular measure used to define the given direct integral, even when not mentioned explicitly.

It has been well known for some time that it is possible to introduce a topology on a measure space. It seems unlikely that the introduction of such a topology into the measure space used for the given direct integral will make it possible to eliminate the "almost all" statement from most of the deeper results on generalized direct sums. Thus we shall not introduce the above mentioned topology in the present paper, but base our assertions and proofs about direct integrals on [7], [8] and [15]. It is however clear that in any particular case where a topology is really wanted, our results and methods can readily be translated. This remark applies in particular to the results outlined in [11a]. There it seems of interest to consider the measure space in question more closely. We plan to come back to this in a later publication, where it will be shown that the methods of the present paper and of [11a] lead to more precise results especially for semi-simple Lie groups and throw some new light also on the problem of eigenfunction expansions for certain partial differential equations (both of the ellipitic and hyperbolic type) especially when there is a continuous spectrum.

2. Isomorphisms of factors. In this section let Ξ be an arbitrary Hilbert space over the complex numbers, and $\boldsymbol{M}$ a factor in Ξ; i. e. $\boldsymbol{M}$ is a weakly closed self-adjoint algebra of bounded operators in Ξ whose center consists of the scalar multiples of the identity operator I. According to [12] there exists on $\boldsymbol{M}$ an essentially unique relative dimension function $d(E)$ defined for all projections $E \varepsilon \boldsymbol{M}$. According to a result of Rickart (Cor. 4.13 of [17]) the projections $E \varepsilon \boldsymbol{M}$ with $d(E) < \infty$ are contained in every proper two-sided ideal of $\boldsymbol{M}$. There exists therefore a two-sided ideal $\boldsymbol{J}$ of $\boldsymbol{M}$ con-

taining all projections of finite relative dimension such that every other non-zero two-sided ideal of $\boldsymbol{M}$ contains $\boldsymbol{J}$. The following lemma is only stated explicitly for the convenience of the reader and is not essentially new.

LEMMA 2. 1. *$\boldsymbol{M}$ is the smallest weakly closed self-adjoint operator algebra which contains the identity operator I and the ideal $\boldsymbol{J}$.*

Proof. If $\boldsymbol{M}$ is of finite type in the sense of [12], then it has no proper two-sided ideals; hence $\boldsymbol{J} = \boldsymbol{M}$ in this case, and hence our Lemma 2. 1 is (trivially) true in this case.

If $\boldsymbol{M}$ is a factor of infinite type, then Theorem VIII of [12] implies that there exists for every integer $n \geq 0$ a projection $E_n \varepsilon \boldsymbol{M}$ with $d(E_n) = n$. It follows from Lemma 8. 13 of [12] that we can assume $E_n < E_{n+1}$. Therefore the E_n converge strongly to a projection $E \varepsilon \boldsymbol{M}$. Since $d(E - E_n) \geq 0$, we have $d(E) = d(E - E_n) + d(E_n) \geq n$ for all n; thus $d(E) = \infty$. Now let F be an arbitrary projection $\boldsymbol{M}$ with $d(F) = \infty$. By Lemma 8. 13 of [12] E and F are equivalent in the sense that there exists a partially isometric operator $\pi \varepsilon \boldsymbol{M}$ such that $E = \pi\pi^*$ and $F = \pi^*\pi$, whence $\pi^* E \pi = F$. Put $F_n = \pi^* E_n \pi$. Then strong $\lim_n F_n = F$ and $d(F_n) = d(F_n)$. So we have found for every projection $F \varepsilon \boldsymbol{M}$ with $d(F) = \infty$ and ascending sequence of projections $\varepsilon \boldsymbol{J}$ converging strongly to F. This together with the fact that $\boldsymbol{M}$ is generated (as a weakly or strongly closed operator algebra) by its projection (cf. [16]) proves Lemma 2. 1.

As an immediate consequence we obtain

LEMMA 2. 2. *Let $\boldsymbol{M}$ be a factor and C an arbitrary idempotent element of $\boldsymbol{M}'$, i. e. $C^2 = C$ is a bounded linear operator, and $CA = AC$ for all $A \varepsilon \boldsymbol{M}$. Then the mapping $A \rightarrow CA$ is an isomorphism of $\boldsymbol{M}$ onto the algebra $C\boldsymbol{M}$, provided $C \neq 0$.*

Proof. The mapping $A \rightarrow CA$ is clearly a homomorphism of the algebra $\boldsymbol{M}$ since $C^2 = C$ and $AC = CA$. Hence the set of elements A of $\boldsymbol{M}$ for which $AC = 0$ is a two-sided ideal of $\boldsymbol{M}$. If it were not the zero ideal it would have to contain $\boldsymbol{J}$, by the above mentioned result of Rickart. But then Lemma 2. 1 above would imply $AC = 0$ for all $A \varepsilon \boldsymbol{M}$, which contradicts $IC = C \neq 0$, since the identity operator I is an element of $\boldsymbol{M}$.

3. Decomposition of an invariant subspace. In this section let $\boldsymbol{H}$ be an arbitrary (separable) Hilbert space and $\boldsymbol{W}$ an arbitrary weakly closed self-adjoint algebra of bounded operators in $\boldsymbol{H}$. Perform the central decomposition (Theorem VII of [15]) under the algebra $\boldsymbol{W}$:

$$(3.1) \qquad \boldsymbol{H} = \int_{\oplus} \boldsymbol{H}_t.$$

Now let $E = E^* = E^2 \,\varepsilon\, \boldsymbol{W}$, where $\boldsymbol{W}'$ denotes as usual the commuting algebra of $\boldsymbol{W}$. By Theorem V of [15] $E \,\varepsilon\, \boldsymbol{W}'$ implies that E *is* decomposable under (3.1) into an operator-valued function, say $E(t)$. Let $\boldsymbol{T}_1$ be the set of those t for which $E(t) \neq 0(t)$, where $0(t)$ denotes the zero-operator in the space $\boldsymbol{H}_t$. Put $E\boldsymbol{H} = \boldsymbol{H}_1$, and $E(t)\boldsymbol{H}_t = \boldsymbol{H}_{1,t}$, for $t \,\varepsilon\, \boldsymbol{T}_1$. Clearly for any $x \,\varepsilon\, \boldsymbol{H}$ we have $x \,\varepsilon\, \boldsymbol{H}_1$ if and only if

$$(3.2) \qquad x(t) = 0(t) \quad \text{for } t \not\varepsilon\, \boldsymbol{T}_1$$
$$\text{and} \quad x(t) \,\varepsilon\, \boldsymbol{H}_{1t} \quad \text{for } t \,\varepsilon\, \boldsymbol{T}_1$$

(after a possible change of $x(t)$ on a t-set of measure zero which we assume to have been made). We have therefore a one-one correspondence between the elements x of $\boldsymbol{H}_1$ and those equivalence classes of vector valued functions $x(t)$ which occur in (3.1) and satisfy the conditions (3.2). It is easy to conclude from this that this one-one correspondence defines a direct integral decomposition

$$(3.3) \qquad \boldsymbol{H}_1 = \int_{\oplus} \boldsymbol{H}_{1t},$$

where t now ranges only over the set $\boldsymbol{T}_1$ and the measure used in the definition of (3.3) is the restriction of the measure used in the definition of (3.1) to the subset $\boldsymbol{T}_1$.

Denote by $\boldsymbol{Z}$ the center of $\boldsymbol{W}$ $(\boldsymbol{Z} = \boldsymbol{W} \cap \boldsymbol{W}')$, by $\boldsymbol{W}_1$ the algebra $E\boldsymbol{W}$, and by $\boldsymbol{Z}_1$ the algebra $E\boldsymbol{Z}$, where the elements of $\boldsymbol{W}_1$ and $\boldsymbol{Z}_1$ are considered to be operators of the space $\boldsymbol{H}_1$ into itself. We then have

LEMMA 3.1. *The direct integral* (3.3) *is the central decomposition of the space* $\boldsymbol{H}_1$ *under the algebra* $\boldsymbol{W}_1$ *i. e. the algebra* $\boldsymbol{Z}_1$ *"belongs to it" in the sense of* [15]. *In particular* $\boldsymbol{Z}_1$ *is the center of* $\boldsymbol{W}_1$.

Proof. Let $A_1 \,\varepsilon\, \boldsymbol{W}_1$; then there exists an element A of $\boldsymbol{W}$ such that A_1 is the restriction of EA to $\boldsymbol{H}_1$. Since $A \,\varepsilon\, \boldsymbol{W}$, it is decomposable under (3.1) into an operator-valued function $A(t)$ say. Moreover $(EA)(t) = E(t)A(t)$ for almost all t implies

$$(3.4) \qquad E(t)A(t)\boldsymbol{H}_{1t} \subseteqq \boldsymbol{H}_{1t}$$

for almost all t. Now change $A(t)$ on a set of measure zero so that (3.4) becomes valid for *all* t and put

$$(3.5) \qquad A_1(t) = \text{restriction of } E(t)A(t) \text{ to } \boldsymbol{H}_{1t} \text{ for } t \,\varepsilon\, \boldsymbol{T}_1.$$

Then A_1 decomposes into $A_1(t)$ under (3.3).

Let us now choose countably many elements $A^{(j)}$ $(j = 1, 2, \cdots)$ of $\boldsymbol{W}$ which generate $\boldsymbol{W}$ (in the weak or strong topology). It follows from the proof of Theorem VI (on p. 459) of [15] that we may assume $\boldsymbol{W}(t)$ to be generated by the elements $A^{(j)}(t)$ for every t where we can also assume that the $A^{(j)}(t)$ are chosen such that (3.4) holds for every t and also

$$(3.6) \qquad A^{(j)}(t)E(t) = E(t)A^{(j)}(t) \text{ for every } t \text{ and all } j = 1, 2, 3, \cdots.$$

Now let $\boldsymbol{W}_1(t)$ be the (weakly closed self-adjoint) algebra generated by the operators $A_1^{(j)}(t)$ for $t \in \boldsymbol{T}_1$. Then it is clear that since $E(t) \in \boldsymbol{W}'(t) = \boldsymbol{W}(t)'$ (cf. Lemma 13 of [15]), the mapping

$$(3.7) \qquad A(t) \to A_1(t) = \text{restriction of } E(t)A(t) \text{ to } \boldsymbol{H}_{1t}$$

is a homomorphism of $\boldsymbol{W}(t)$ onto $\boldsymbol{W}_1(t)$ for almost every $t \in \boldsymbol{T}_1$. Since $E(t) \neq 0(t)$ for $t \in \boldsymbol{T}_1$, and $\boldsymbol{W}(t)$ is a factor for every t, we may apply Lemma 2.2 above and conclude that *the mapping* (3.7) *is an isomorphism of* $W(t)$ *onto* $W_1(t)$ *for almost every* $t \in \boldsymbol{T}_1$. Hence in particular $\boldsymbol{W}_1(t)$ is also a factor for almost every $t \in \boldsymbol{T}_1$.

Note that it follows from the way $\boldsymbol{W}_1(t)$ is defined that if X is an arbitrary element of $\boldsymbol{W}_1$, i. e. $X = AE = EA$ with $A \in \boldsymbol{W}$, then $X(t)$ = above $A_1(t)$ is an element of $\boldsymbol{W}_1(t)$ for almost every $t \in \boldsymbol{T}_1$. Observe also that $\boldsymbol{W}_1(t)$ depends measurably on t for $t \in \boldsymbol{T}_1$ in the sense of definition 5 of [15] under the direct integral (3.3). Hence in order to prove that (3.3) is the central decomposition under $\boldsymbol{W}_1$ it is by Theorem VII of [15] sufficient to prove the following: If $Y(t)$ is an arbitrary bounded measurable operator valued function (in the sense of Definition 5 of [15]) defined for all $t \in \boldsymbol{T}_1$, and satisfies $Y(t) \in \boldsymbol{W}_1(t)$, then there exists an operator Y of the space $\boldsymbol{H}_1$ which is an element of $\boldsymbol{W}_1$ and decomposes into $Y(t)$ under (3.3); i. e. we have to prove $\boldsymbol{W}_1 \approx \sum \boldsymbol{W}_1(t)$ in the terminology of [15] under the direct integral (3.3).

By Lemma 12 of [15] the ring generated by $\boldsymbol{P}_1$ and the operators $A_1^{(j)}$ satisfies $\approx \sum \boldsymbol{W}_1(t)$, where $\boldsymbol{P}_1$ denotes the ring of those operators of $\boldsymbol{H}_1$ which decomposes into scalars under (3.3). Since the $A_1^{(j)}$ generate $\boldsymbol{W}_1$ it remains to prove $\boldsymbol{P}_1 \subseteqq \boldsymbol{W}_1$.

$C_1 \in \boldsymbol{P}_1$ means that there exists a complex valued bounded measurable function $c(t)$ defined for $t \in \boldsymbol{T}_1$, such that C_1 decomposes under (3.3) into $C_1(t) = c(t)I_1(t)$, where $I_1(t)$ denotes the identity operator in $\boldsymbol{H}_{1t}$. Let $C(t)$denote $c(t)I(t)$ or $0(t)$ according as t is or is not in $\boldsymbol{T}_1$. Then

$$(3.8) \qquad C_1(t) = \text{restriction of } C(t)E(t) \text{ to } \boldsymbol{H}_{1t} \text{ for } t \in \boldsymbol{T}_1,$$

because $I_1(t)$ = restriction of $E(t)$ to $\boldsymbol{H}_{1t}$ for $t \in \boldsymbol{T}_1$. Since $\boldsymbol{T}_1$ is a measurable set and $c(t)$ a measurable function, the operator-valued function $C(t)$ depends measurably on t. Hence there exists an operator C in $\boldsymbol{H}$ which decomposes into $C(t)$ under (3.1). Since $C(t)$ is a scalar for every t, and since (3.1) belongs by hypothesis to $\boldsymbol{Z}$ (which means that $\boldsymbol{Z}$ is exactly the ring of those operators in $\boldsymbol{H}$ which decompose into scalars under (3.11)), we have $C \in \boldsymbol{Z}$. But $Z \subseteqq \boldsymbol{W}$, and hence $C \in \boldsymbol{W}$. Also (3.8) implies that C_1 is the restriction of CE to $\boldsymbol{H}_1$. This proves $C_1 \in \boldsymbol{W}_1$, i. e. $\boldsymbol{P}_1 \subseteqq \boldsymbol{W}_1$. So we have proved $\boldsymbol{W}_1 \approx \sum \boldsymbol{W}_1(t)$ in the terminology of [15]. As remarked above, this fact proves that (3.3) is the central decomposition of $\boldsymbol{W}_1$. This proves the first assertion of Lemma 3.1.

But the fact (which we have just proved) that (3.3) is the central decomposition of $\boldsymbol{H}_1$ under $\boldsymbol{W}_1$ implies that $\boldsymbol{P}_1$ is the center of $\boldsymbol{W}_1$. On the other hand we have just seen that (3.8) implies that the elements C_1 of $\boldsymbol{P}_1$ are exactly the elements of the form "restriction of CE to $\boldsymbol{H}_1$" where C varies over $\boldsymbol{Z}$, i. e. the elements of $\boldsymbol{Z}_1$. This proves $\boldsymbol{Z}_1 = \boldsymbol{P}_1$; hence $\boldsymbol{Z}_1$ is the center of $\boldsymbol{W}_1$. This proves the second assertion of Lemma 3.1 and hence the proof of Lemma 3.1 is complete.

Later on we shall also require

LEMMA 3.2. *Let again $E = E^* = E^2$ and $E \in \boldsymbol{W}'$. Let C_E be the smallest projection $\in \boldsymbol{Z} = \boldsymbol{W} \cap \boldsymbol{W}'$ which satisfies $C_E \supseteqq E$. Then the restriction of the homomorphism $X \to XE$ to the subalgebra $\boldsymbol{W}C_E$ of $\boldsymbol{W}$ is an isomorphism of $\boldsymbol{W}C_E$ onto $\boldsymbol{W}E$. The kernel of the homomorphism $X \to XE$ is $\boldsymbol{W}(I - C_E)$.*

Proof. By hypothesis $C_E \supseteqq E$, i. e. $C_E E = E$; hence $\boldsymbol{W}E = (\boldsymbol{W}C_E)E$, which proves that the onto-assertion of the lemma is trivial. In order to prove that $\boldsymbol{W}C_E$ is mapped isomorphically under the homomorphism

$$X \to XE, \tag{3.9}$$

it is clearly sufficient to prove that $\boldsymbol{W}(I - C_E)$ is the kernel of (3.9). Let $C_E(t)$ be the operator valued function into which C_E decomposes under (3.1). Since $C_E \in \boldsymbol{Z}$, and since $\boldsymbol{Z}$ belongs to (3.1), we have as in the proof of Lemma 3.1 $C_E(t) = c(t)I(t)$, where $c(t)$ is a numerical essentially bounded measurable function defined for all t. Since C_E is a projection, we must have $c(t) = 1$ or 0 for almost all t. But $C_E E = E$ implies $C_E(t)E(t) = E(t)$ for almost all t. Since $E(t) \neq 0(t)$ for $t \in \boldsymbol{T}_1$, we must have $c(t) = 1$ for almost all $t \in T_1$. Since $E(t) = 0(t)$ for $t \notin T_1$, we must have $c(t) = 0$ for almost all $t \notin \boldsymbol{T}_1$, for otherwise C_E would not be minimal among the central projections $\supseteqq E$. This proves

$$(3.10) \qquad C_E(t) = \begin{cases} I(t) & \text{for almost all } t \in \boldsymbol{T}_1 \\ 0(t) & \text{for almost all } t \notin \boldsymbol{T}_1. \end{cases}$$

Now suppose X is in the kernel of the homomorphism (3.9), i. e. suppose $XE = 0$. This means $X(t)E(t) = 0(t)$ for almost all t under the decomposition (3.1). If $t \in \boldsymbol{T}_1$, then Lemma 2.2 implies for almost all t that $X(t)E(t) = 0(t)$ if and only if $X(t) = 0(t)$. If on the other hand $t \notin \boldsymbol{T}_1$, then $X(t)E(t) = 0(t)$ for *any* $X \in \boldsymbol{W}$. This proves that $XE = 0$ if and only if $X(t) = 0(t)$ for almost all $t \notin \boldsymbol{T}_1$. But by (3.10) this last statement is the same as $XC_E = 0$, i. e. $X(I - C_E) = I$, which proves that the kernel of the homomorphism (3.9) is exactly $\boldsymbol{W}(I - C_E)$.

In the course of the proof of Lemma 3.2 we obtained the following

COROLLARY 3.1. *Let C_E be the smallest projection $\varepsilon\, \boldsymbol{Z}$ which satisfies $C_E \supseteqq E$. Let $C_E(t)$ be the operator valued function into which C_E decomposes under the central decomposition* (3.1). *Then $C_E(t)$ is given for almost all t by* (3.10), *where $\boldsymbol{T}_1$ is again the set of these t for which $E(t) \neq 0(t)$.*

4. The central decomposition of the regular representation of G. In this section we consider the central decomposition of the regular representation of our separable locally compact unimodular group G. Denote by $\boldsymbol{R}$ the weakly closed self-adjoint operator algebra generated by the right translations $R(g)$ in $\mathfrak{L}_2(G)$, and by $\boldsymbol{L}$ the algebra generated by the left translations $L(g)$. Godement and Segal have shown that $\boldsymbol{L}$ and $\boldsymbol{R}$ are each other's commuting algebras:

$$(4.1) \qquad \boldsymbol{L} = \boldsymbol{R}' \quad \text{and} \quad \boldsymbol{R} = \boldsymbol{L}'.$$

Denote by $\boldsymbol{Z}$ the center of $\boldsymbol{R}$: $\boldsymbol{Z} = \boldsymbol{R} \cap \boldsymbol{R}' = \boldsymbol{R} \cap \boldsymbol{L} = \boldsymbol{L} \cap \boldsymbol{L}'$, and let

$$(4.2) \qquad \mathfrak{L}_2(G) = \int_{\oplus} \mathfrak{H}_t$$

be the central decomposition of $\mathfrak{L}_2(G)$ under $\boldsymbol{R}$ (or $\boldsymbol{L}$).

If $a(g)$ is an arbitrary complex valued Haar-Lebesgue-integrable function on G, let $\boldsymbol{R}_a$ be the operator acting on $\mathfrak{L}_2(G)$ defined by

$$(4.3) \qquad R_a = \int_G a(g) R(g)\, dg.$$

Clearly R_a is an element of $\boldsymbol{R}$. Denote by $\boldsymbol{R}^{(1)}$ the subset of $\boldsymbol{R}$ of the elements R_a obtained from integrable functions $a(g)$, and by $\boldsymbol{R}^{(1,2)}$ the subset of those $R_a \in \boldsymbol{R}^{(1)}$ for which $a(g)$ is also square integrable. Under the decomposition (4.2) there corresponds to the operator R_a an operator-valued function, say

$A(t)$, and to the element $a(g)$ of $\mathfrak{L}_2(G)$ a vector valued function $a(t)$, whenever $a(g) \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$. Again $A(t)$ and $a(t)$ can be changed arbitrarily on sets of measure zero. However we have

LEMMA 4.1. *Let $a(g) \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$. Then there exists one choice which assigns to every operator R_a a unique operator valued function $A(t)$ and to the "vector" $a(g)$ a unique vector valued function $a(t)$ such that for this choice the mapping*

$$A(t) \rightarrow a(t) \tag{4.4}$$

is for almost every t a one-one linear mapping of a certain weakly dense linear subspace $\boldsymbol{P}(t)$ of the operator algebra $\boldsymbol{R}(t)$ defined below, onto a dense linear subspace of $\mathfrak{H}_t$.

Proof. It is clear that $\boldsymbol{R}$ is generated by $\boldsymbol{R}^{(1,2)}$. It follows from p. 386 of [16] that there exists a sequence of functions $a_j(g) \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$ $(j = 1, 2, \cdots)$ which are dense in $\mathfrak{L}_2(G)$ and for which the corresponding operators

$$R_{a_1}, R_{a_2}, \cdots \tag{4.4a}$$

generate the ring $\boldsymbol{R}$ as the smallest weakly closed self-adjoint operator algebra containing the operators (4.4a). Now choose for each operator in the sequence (4.4a) *one* operator valued function $A_j(t)$ into which it decomposes under our given direct integral. Then it follows from the proof of Theorem VI (on p. 459) of [15] that after one possible change on a set of measure zero the factors $\boldsymbol{R}(t)$ into which $\boldsymbol{R}$ decomposes may be assumed to be generated by the operators $A_1(t), A_2(t), \cdots$. Now let A be a (not necessarily commutative) polynomial

$$p(R_{a_{j_1}}, R_{a_{j_2}}, \cdots) \tag{4.4b}$$

in a finite number of the operators (4.4a), and define $A(t)$ to be equal to

$$p(A_{j_1}(t), A_{j_2}(t), \cdots). \tag{4.4c}$$

Denote the family of operators so obtained for each t by $\boldsymbol{P}(t)$. Then $\boldsymbol{P}(t)$ is a dense linear subspace of $\boldsymbol{R}(t)$ in the weak topology for operators.

Now consider all monomials in some of the operators (4.4a). Since to the product $R_a R_{a'}$ corresponds the convolution of the functions $a(g)$ and $a'(g)$, i.e. $R_a R_a = R_{a''}$ where $a''(g) = \int a(g\gamma^{-1}) a'(\gamma) d\gamma$ and where $a'' \varepsilon \mathfrak{L}_1(G)$ whenever a and $a' \varepsilon \mathfrak{L}_1(G)$ it follows (for instance from Lemma 7.1 of [8]) that $a \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$ and $a' \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$ imply $a'' \varepsilon \mathfrak{L}_1(G) \cap \mathfrak{L}_2(G)$. Hence there exists a sequence of functions $b_1(g), b_2(g), \cdots$ each $\varepsilon \mathfrak{L}_1(G)$

$\cap \mathfrak{L}_2(G)$ such that each of the above monomials is equal to one of the operators R_{b_m}.

Clearly the functions $b_m(g)$ are dense in $\mathfrak{L}_2(G)$ since the $a_j(g)$ are contained among the $b_m(g)$. Now choose for each $b_m(g)$ one vector valued function $b_m(t)$ into which it decomposes under (4.2). It is proved in § 6 of [15] that the finite linear combinations of the $b_m(t)$ form a dense linear subspace of $\mathfrak{H}_t$ for almost every t and hence, after one change on a set of measure zero, for all t. Now let $x(g)$ be an arbitrary finite linear combination of the $b_m(g)$ with complex coefficients c_{m_i}:

$$x(g) = \sum_{i=1}^{r} c_{m_i} b_{m_i}(g) \qquad (r \text{ arbitrary } < \infty). \tag{4.5}$$

Define the vector valued function $x(t)$ by

$$x(t) = \sum_{i=1}^{r} c_{m_i} b_{m_i}(t) \tag{4.6}$$

then $x(g)$ decomposes into $x(t)$ under (4.2). As remarked above the $x(t)$ form for each fixed t a dense linear subspace of $\mathfrak{H}_t$, and the operators $X(t)$ defined by

$$X(t) = \sum_{i=1}^{r} c_{m_i} B_{m_i}(t) \tag{4.7}$$

are exactly the polynomials $p(A_{j_1}(t), A_{j_2}(t), \cdots)$ defined above. Hence the $X(t)$ form the dense subalgebra $\boldsymbol{P}(t)$ of $\boldsymbol{R}(t)$. Therefore the proof of Lemma 4.1 will be complete if we prove the following:

The correspondence $X(t) \leftrightarrow x(t)$ between the elements (4.7) *of $\boldsymbol{P}(t)$ and the elements* (4.6) *of $\mathfrak{H}_t$ is a one-one linear mapping for almost every t.* Denote by T_t the (essentially unique) relative trace which exists for certain elements of the factor $\boldsymbol{R}(t)$ in accordance with [13]. According to Lemma 7.4 of [8] the factor $\boldsymbol{R}(t)$ will be of type I or II for every t (after a possible change on one set of measure zero which we assume to have been made). Then there exists a certain function $\alpha(t) > 0$ and $\leq \infty$ such that

$$\int x(g)\bar{y}(g)dg = \int \alpha(t) T_t(X(t)Y(t)^*)ds(t), \tag{4.8}$$

where $ds(t)$ refers to the measure used to define the direct integral (4.2). On the other hand $\int x(g)\bar{y}(g)dg = \int (x(t), y(t))ds(t)$. Now let C be any element of $\boldsymbol{R}$, it decomposes into a certain operator valued function, say $C(t)$, under (4.2), and we obtain

$$\begin{aligned} \int (Cx)(g)y(g)dg &= \int (C(t)x(t), y(t))ds(t) \\ &= \int \alpha(t) T_t(C(t)X(t)Y(t)^*)ds(t). \end{aligned} \tag{4.9}$$

In accordance with Theorem IV and VII of [15] we obtain, as C ranges over the center $\boldsymbol{Z}$ of $\boldsymbol{R}$, for the functions $C(t)$ *all* essentially bounded *scalar valued* functions $c(t)I(t)$. Hence for $C \varepsilon \boldsymbol{Z}$ we obtain

$$\int (c(t)x(t), y(t))ds(t) = \int \alpha(t)c(t)T_t(X(t)Y(t)^*)ds(t).$$

In this equation $c(t)$ can be the characteristic function of an arbitrary s-measurable set, which proves

$$(x(t), y(t)) = \alpha(t)T_t(X(t)Y(t)^*) \tag{4.10}$$

for all t outside of some set of measure zero which may depend on $x(g)$ and $y(g)$. However if we put $b_m(g)$ for $x(g)$ and $b_n(g)$ for $y(g)$, we get from (4.10)

$$(b_m(t), b_n(t)) = \alpha(t)T_t(B_m(t)B_n(t)^*), \tag{4.11}$$

for all t outside of one set of measure zero which is obtained by taking the union of the countably many sets corresponding to each pair m, n of integers. Hence if $x(g)$ and $y(g)$ are finite linear combinations of the functions $b_m(g)$, then we can define $X(t)$ and $Y(t)$ in terms of the functions $B_m(t)$ by equation (4.7), and $x(t)$ and $y(t)$ by equation (4.6), and obtain for this particular choice of the functions $x(t), y(t), X(t), Y(t)$ the truth of (4.10) for all t outside of one sea of measure zero.

It is proved in Theorem VIII of [15] that the above function $\alpha(t)$ (denoted there by $a(\lambda)$) is positive for all t. If we can prove that $\alpha(t)$ is finite for almost all t then we may consider the right side of equation (4.10) above to be an inner product defined on $\boldsymbol{P}(t)$. The left side of (4.10) is the inner product of the space $\mathfrak{H}_t$. Hence $\alpha(t) < \infty$ would imply that the correspondence $X(t) \leftrightarrow x(t)$ preserves inner products, hence is one-one and clearly linear for almost every t (and therefore for all t, after a change on one set of measure zero). Therefore the proof of Lemma 4.1 will be complete if we prove

LEMMA 4.2. *The function $\alpha(t)$ which occurs in the generalized Peter-Weyl-Plancherel formula* (4.8) *is finite and positive for almost every t.*

Proof. As remarked above, $\alpha(t) > 0$ was proved by von Neumann ([15], Theorem VIII). To prove $\alpha(t) < \infty$ we observe that for $m = n$ we obtain from (4.11)

$$\| b_n(t) \|^2 = \alpha(t)T_t(B_n(t)B_n(t)^*).$$

Since $\| b_n(t) \|^2 < \infty$, the expression $T_t(B_n(t)B_n(t)^*)$ would have to be 0 for any t for which $\alpha(t) = \infty$. But then

$$B_n(t) = 0(t). \tag{4.12}$$

The set of those t for which (4.12) holds is known to be measurable (cf. [15], § 13); hence so is the set of those t for which (4.12) holds for all $n = 1, 2, \cdots$. On the other hand the $B_n(t)$ generate the ring $\boldsymbol{R}(t)$, and $\boldsymbol{R}(t)$ contains the identity operator $I(t)$ for almost every t. Hence (4.12) can be true only on a t-set of measure zero. Therefore $\alpha(t) < \infty$ for almost every t as required. This completes the proof of Lemma 4.2 and hence also the proof of Lemma 4.1.

In the next section we shall require a special case of the following

LEMMA 4.3. *Let β be a countably additive complex valued set function on G. For $y(g)$, a suitable complex valued function on G, put*

$$(4.12) \qquad (R_\beta y)(g) = \int_G y(g\gamma^{-1})d\beta(\gamma) \text{ and } (L_\beta y)(g) = \int_G y(\gamma^{-1}g)d\beta(\gamma).$$

Assume β is such that R_β and L_β are bounded linear operators on $\mathfrak{L}_2(G)$ defined for all $y(g) \varepsilon \mathfrak{L}_2(G)$ and that R_β and L_β are defined for all $y(g) \varepsilon \mathfrak{L}_1(G)$ and satisfy $R_\beta\mathfrak{L}_1(G) \subset \mathfrak{L}_1(G)$ and $L_\beta\mathfrak{L}_1(G) \subset \mathfrak{L}_1(G)$. Denote by $R_\beta(t)$, $L_\beta(t)$ the operator valued functions into which the operators R_β, L_β respectively decompose under the central decomposition (4.2).

ASSERTION. *The mapping* (4.4) *can be taken to be such that under it there corresponds for almost every fixed t to the operator $A(t)R_\beta(t)$ the element $L_\beta(t)a(t)$ of $\mathfrak{H}_t$, and to the operator $R_\beta(t)A(t)$ the element $R_\beta(t)a(t)$ of $\mathfrak{H}_t$.*

Proof. Observe first that the definition (4.12) of R_β, together with the assumption that R_β be a bounded operator, implies $R_\beta \varepsilon \boldsymbol{R}$; similarly $L_\beta \varepsilon \boldsymbol{L}$. Hence R_β and L_β are decomposable operators under (4.2). Therefore the above function $R_\beta(t)$ and $L_\beta(t)$ exist.

Notice next that the definition (4.12) implies that

$$(4.13) \qquad R_a R_\beta = R_{L_\beta a} \quad \text{and} \quad R_\beta R_a = R_{R_\beta a}.$$

Hence for all t outside of a set of measure zero (which may depend on the function $a(g)$ and the measure β) we obtain

$$(4.14) \qquad R_a(t)R_\beta(t) = R_{L_\beta a}(t), \text{ and } R_\beta(t)R_a(t) = R_{R_\beta a}(t),$$

where $R_a(t), \cdots$ are arbitrary operator valued functions into which the operators $R_a, \cdots$ decpompose under (4.2) (arbitrary in the sense that they may be changed arbitrarily on sets of measure zero). Now put in equations (4.14) the function $b_n(g)$ instead of $a(g)$ and replace the "arbitrary" $R_a(t)$ by the function $A(t)$ introduced in the proof of Lemma 4.1. By taking the union of countably many sets of measure zero we then obtain from (4.14)

$$A(t)R_\beta(t) = R_{L_\beta a}(t), \text{ and } R_\beta(t)A(t) = R_{R_\beta a}(t),$$

for all t outside of *one* set of measure zero whenever $a(g)$ is a finite linear combination of the functions $b_n(g)$. By hypothesis the functions $(L_\beta a)(g)$ and $(R_\beta a)(g)$ are elements of $L_1(G) \cap L_2(G)$ whenever $a(g) \varepsilon L_1(G) \cap L_2(G)$. Since the measure β is fixed, we may assume that the functions $b_n(g)$ are chosen such that $L_\beta b_n = b_{n'}$ and $R_\beta b_n = b_{n''}$ are again functions of the sequence $b_1(g), b_2(g), \cdots$. Put $(L_\beta a)(g) = {}_\beta a(g)$ and $(R_\beta a)(g) = a_\beta(g)$, and let $A_\beta(t)$, ${}_\beta A(t)$ be the operator-valued function of Lemma 4.1 into which $R_{R_\beta a}$, $R_{L_\beta a}$ decompose respectively. Then we get all t outside of *one* set of measure zero

$$(4.15) \qquad A(t)R_\beta(t) = {}_\beta A(t), \text{ and } R_\beta(t)A(t) = A_\beta(t).$$

On the other hand $(L_\beta a)(g)$ decomposes into the vector valued function $(L_\beta a)(t) = {}_\beta a(t)$ say, for almost all t, and $R_\beta a$ into $A_\beta(t)a(t) = a_\beta(t)$ for almost all t, where ${}_\beta a(t)$ and $a_\beta(t)$ are the vector valued functions uniquely determined above by $a(g)$ for which Lemma 4.1 is true.

Moreover, since we may assume $L_\beta b_n = b_{n'}$ and $R_\beta b_n = b_{n''}$ we see that the operators ${}_\beta A(t)$ and $A_\beta(t)$ are elements of $P(t)$; therefore the mapping (4.4) is defined for ${}_\beta A(t)$ and $A_\beta(t)$, and for all t outside of *one* set of measure zero we obtain

$${}_\beta A(t) \to {}_\beta a(t) \text{ and } A_\beta(t) \to a_\beta(t),$$

under the mapping (4.4). Combining this with (4.15) we obtain the truth of Lemma 4.3.

5. Completion of the proof of Theorem 1. Let us now consider the repersentation $U = \operatorname{ind} u$ of G defined in **1**. If the representation $u\colon k \to u(k)$ of K is a discrete direct sum of representations u_ν, $u(k) = \sum_\oplus u_\nu(k)$, then it follows that U is the direct sum of representation U_ν of G, where $U_\nu = \operatorname{ind} u_\nu$:

$$(5.1) \qquad \operatorname{ind}\{\sum_\oplus {}_\nu u_\nu\} = \sum_\oplus {}_\nu \operatorname{ind} u_\nu.$$

Indeed if the representation space $\mathfrak{h}$ of u is a direct sum $\sum_\oplus \mathfrak{h}_\nu$ of invariant subspace $\mathfrak{h}_\nu$, then consider those functions $X(g) \varepsilon \mathfrak{H}$ for which $X(g)$ is, for each fixed $g \varepsilon G$, an element of $\mathfrak{h}_\nu$. They form a closed invariant linear subspace $\mathfrak{H}_\nu$ of $\mathfrak{H}$ under the operators $U(g)$, and if $U_\nu(g)$ denotes the restriction of $U(g)$ to $\mathfrak{H}$, then the definition of induced representation as given in **1** implies that U_ν is (up to unitary equivalence) equal to $\operatorname{ind} u_\nu$. This proves (5.1).

In particular let λ be the left-regular representation of K. Then

$\lambda = \sum_{\oplus} h_\nu u_\nu$, where u_ν varies over all irreducible representations of the compact group K (equivalent representations being identified); u_ν is not equivalent to $u_{\nu'}$ for $\nu \neq \nu'$, h_ν denotes the degree of u_ν and $h_\nu u_\nu$ denotes the repetition of the representation u_ν, h_ν times. Hence (5.1) gives

$$\text{(5.2)} \qquad \operatorname*{ind}_{K\uparrow G} \lambda = \sum_{\oplus} h_\nu \operatorname{ind} u_\nu.$$

On the other hand the definition of induced repersentation implies that if λ is the regular representation of K (i. e. $\lambda(k)$ is left translation by the element k in the space $\mathfrak{L}_2(K)$), then $L = \operatorname{ind} \lambda$ is the (left-) regular representation of G. To see this, observe that the definition of induced representation implies that the repersentation space $\mathfrak{H}_\lambda$ for $\operatorname{ind} \lambda$ can be identified with the space of those elements $x(k, g)$ of $\mathfrak{L}_2(K) \times \mathfrak{L}_2(G)$ which satisfy

$$\text{(5.3)} \qquad x(k, k'g) = x(kk'^{-1}, g).$$

Hence the square norm equals

$$\int_K \int_G |x(k, g)|^2 dg dk = \int_K \{ \int_G |x(1, k^{-1}g)|^2 dg\} dk$$
$$= \int_K \{ \int_G |x(1, g)|^2 dg\} dk = \int_G |x(1, g)|^2 dg,$$

where we take the Haar measure dk on K to be normalized so that $\int_K dk = 1$. This proves that if $x(k, g) \,\varepsilon\, \mathfrak{H}_\lambda$, then the mapping

$$\text{(5.4)} \qquad x(k, g) \to x(1, g) = x'(g)$$

is a unitary mapping of $\mathfrak{H}_\lambda$ onto $\mathfrak{L}_2(G)$. Also the mapping (5.4) clearly commutes with left translations by elements of G, which proves that $\operatorname{ind} \lambda$ is (up to unitary equivalence) equal to the left regular reprseentation L of G. The reader who is worried about the fact that the functions $x(k, g)$ are, as elements of $\mathfrak{L}_2(K) \times \mathfrak{L}.(G)$, only defined up to sets of measure zero, may restrict himself at first to functions $x(k, g)$ which are continuous in k, and then extend the isometry (5.4) uniquely to a unitary mapping from $\mathfrak{H}_\lambda$ onto $\mathfrak{L}_2(G)$.

Thus (5.2) can be written as

$$\text{(5.5)} \qquad L = \sum_{\oplus} h_\nu U_\nu, \text{ where } U_\nu = \operatorname{ind} u_\nu.$$

Let χ_ν be the character of u_ν, i. e. $\chi_\nu(k) =$ trace of $u_\nu(k)$. Then the operator

$$\Lambda_{\chi^\nu} = h_\nu \int_K \chi_\nu(k) \lambda(k)\, dk = h_\nu \int_K \chi_\nu(k) \rho(k)\, dk$$

is a projection; here $\rho(k)$ denotes right-translation by the element k of K in the space $\mathfrak{L}_2(K)$. The subspace $\Lambda_{\chi_\nu} \mathfrak{L}_2(K)$ of $\mathfrak{L}_2(K)$ is representation space for the representation $h_\nu u_\nu$ of K, as is well known.

Therefore the space of those elements $x(k,g)$ of $\mathfrak{L}_1(K)\times\mathfrak{L}_2(G)$ which satisfy (5.3) and $h_\nu\int\chi_\nu(k)x(k'k^{-1},g)\,dk=x(k',g)$ can be taken to be representation space for ind $h_\nu u_\nu$ under right translations by the elements of G. Under the unitary mapping (5.4) this space is mapped onto the subspace of all those elements $x'(g)$ of $\mathfrak{L}_2(G)$ which satisfy

$$(5.6)\qquad h_\nu\int_K\chi_\nu(k)x'(gk^{-1})\,dk=x'(g),$$

$$\text{i.e., } (R_{\chi_\nu}x')(g)=x'(g), \text{ if } R_{\chi_\nu}=h_\nu\int_K\chi_\nu(k)R(k)\,dk.$$

This proves

LEMMA 5.1. *Let u be any continuous irreducible (unitary) representation of degree h of the compact subgroup K and χ the character of u. Then the subspace $R_\chi\mathfrak{L}_2(G)$ is a direct sum of h subspaces each of which transforms under left translation by elements of G equivalently to the operators of the representation $U=\operatorname{ind}u$ of G.*[3]

Indeed, since K is compact, any continuous irreducible unitary representation u of K is equivalent to one of the above u_ν, as we have observed above.

Let us now keep the irreducible representation u of K with character χ fixed. Perform as in **4** the central decomposition (4.2) of the space $\mathfrak{L}_2(G)$ under the operators $R(g)$, $g\in G$. The operator R_χ defined by the last of equations (5.6) is a projection which commutes clearly with $L(g)$ for every $g\in G$, hence also with every element of the ring $\boldsymbol{L}$ which the operators $L(g)$ generate. Hence we may apply Lemma 3.1 and conclude that if t is an element of a certain measurable set (of positive measure) $\boldsymbol{T}_1$, then the space $R_\chi(t)\mathfrak{H}_t$ may be identified with the space $\mathfrak{H}_{1t}$ obtained by performing the central decomposition of the space $R_\chi\mathfrak{L}_2(G)$ under the operators $R_\chi L(G)$ of the representation $hU=\operatorname{ind}(hu)$. Here $\boldsymbol{T}_1$ is the set of those t for which $R_\chi(t)\neq 0(t)$.

Put

$$(5.7)\qquad L_\chi=h\int_K\chi(k)L(k)\,dk,$$

and denote by $L_\chi(t)$ the operator-valued function into which the operator L_χ decomposes under the central decomposition (4.2).

---

[3] Note that it follows from this together with Lemma 7.1 of [8] that factors of type III cannot occur (except on a set of measure zero) in the central decomposition of ind u, whenever K is compact and G unimodular and separable.
$K\uparrow G$

LEMMA 5. 2. *Denote by* $\mathbf{R}(t)$ *the factor into which the ring* $\mathbf{R}$ *(generated by the right translation* $R(g)$*) decomposes for almost every* t *under the central decomposition* (4. 2). *Then the restriction of the mapping* $X(t) \to x(t)$ *defined by* (4. 4) *to those elements* $X(t)$ *which satisfy* $X(t) = R_\chi(t)X(t)R_\chi(t)$ *is for almost every* t *a one-one linear mapping between a dense linear subspace of* $R_\chi(t)\mathbf{R}(t)R_\chi(t)$ *and a dense linear subspace of the space* $R_\chi(t)L_\chi(t)\mathfrak{H}_t$. *Hence in particular*

$$\dim R_\chi(t)\mathbf{R}(t)R_\chi(t) = \dim R_\chi(t)L_\chi(t)\mathfrak{H}_t \tag{5.8}$$

where dim *denotes the (ordinary) dimension of a vector space over the field of complex numbers.*

Proof. Replace the measure β of Lemma 4. 3 by the measure $h_\chi(k)\,dk$, where dk refers—as throughout—to the Haar-measure on K. Then the hypotheses of Lemma 4. 3 are readily seen to be verfied by the operators R_χ and L_χ. Hence if $\mathbf{P}(t)$ denotes again the dense subalgebra of $\mathbf{R}(t)$ introduced in Lemma 4. 1, then Lemma 4. 3 tells us that under the mapping (4. 4) the elements of $\mathbf{P}(t)R_\chi(t)$ are mapped into $L_\chi(t)\mathfrak{H}_t$, and the elements of $R_\chi(t)\mathbf{P}(t)$ into $R_\chi(t)\mathfrak{H}_t$. Hence $R_\chi(t)\mathbf{P}(t)R_\chi(t)$ is mapped into the subspace $R_\chi(t)L_\chi(t)\mathfrak{H}_t$ of $\mathfrak{H}_t$ (for almost every t). Denote by $\mathfrak{H}_t^0$ the image of $\mathbf{P}(t)$ (in $\mathfrak{H}_t$) under the mapping (4. 4). Then $\mathfrak{H}_t^0$ is a dense subspace of $\mathfrak{H}_t$ and the image of $R_\chi(t)\mathbf{P}(t)R_\chi(t)$ under the mapping (4. 4) consists—as we have just seen—of $R_\chi(t)L_\chi(t)\mathfrak{H}_t^0$. Hence clearly $R_\chi(t)\mathbf{P}(t)R_\chi(t)$ is a dense subalgebra (in the weak topology for operators) of $R_\chi(t)\mathbf{R}(t)R_\chi(t)$, and $R_\chi(t)L_\chi(t)\mathfrak{H}_t^0$ a dense linear subspace of $R_\chi(t)L_\chi(t)\mathfrak{H}_t$, which proves Lemma 5. 2.

Let us now consider the space $R_\chi(t)\mathfrak{H}_t$. It is an invariant subspace of $\mathfrak{H}_t$ under the elements of the algebra $\mathfrak{L}(t)$ as follows from Lemma 13 of [15]. By Theorem 1. 1 of [8] there exists a continuous unitary representation V_t of G, $V_t\colon g \to V_t(g)$, where the operators act in the space $\mathfrak{H}_t$ and generate the ring $\mathfrak{L}(t)$ according to Lemma 1. 2 of [8]. So we have

$$R_\chi(t)V_t(g) = V_t(g)R_\chi(t)$$

for all g and all t outside of one set of measure zero which is independent of g. Therefore the operators $R_\chi(t)V_t(g)$ form a representation of G in the space $R_\chi(t)\mathfrak{H}_t$ for almost every t.

We shall now prove that

$$L_\chi(t) = h\int_K \chi(k)V_t(k)\,dk \tag{5.9}$$

for almost every t. Let $y_n(g)$ be a sequence of elements of $\mathfrak{L}_1(G)$ such that for the corresponding operators L_{y_n} acting on $\mathfrak{L}_2(G)$ we have strong $\lim L_{y_n} = I$.

Denote by $L_{y_n}(t)$ the operator valued function into which L_{y_n} decomposes under the central decomposition (4.2). In accordance with p. 442 of [15] there exist a subsequence of the y_n for which the $L_{y_n}(t)$ converge strongly to the identity $I(t)$ for almost every t. Let us assume that this subsequence has been chosen and denote it again by y_n. Since $L_{y_n} = \int_G y_n(g)L(g)dg$ by definition, we have for almost every t

$$L_{y_n}(t) = \int_G y_n(g)U(g,t)dg.$$

On the other hand it has been shown in § 1 of [8] that

$$L_{y_n}(t) = \int_G y_n(g)V_t(g)dg$$

for almost every t. Now put $z_n(g) = h\int_K y_n(gk^{-1})\chi(k)dk$; then

$$\text{strong lim } L_{z_n} = \text{strong lim } L_{y_n}L_\chi = L_\chi,$$

since $L_{z_n} = L_{y_n}L_\chi$. Also $L_{z_n}(t) = L_{y_n}(t)L_\chi(t)$ for almost every t implies

$$\begin{aligned}\text{strong}\lim_{n\to\infty} L_{z_n}(t) &= \text{strong}\lim_{n\to\infty}\int_G z_n(g)U(g,t)dg\\ &= \text{strong}\lim_{n\to\infty}\int_G z_n(g)V_t(g)dg = L_\chi(t) = \int_K h\chi(k)U(k,t)dk.\end{aligned}$$

But since the operators $V_t(g)$ form a representation of G, the definition of $z_n(g)$ implies

$$\int z_n(g)V_t(g)dg \to L_\chi(t)h\int\chi(k)V_t(k)dk.$$

Hence strong lim $L_{y_n}(t) = I(t)$ for almost every t implies

$$\text{strong lim }\int z_n(g)V_t(g) = L_\chi(t).$$

for almost every t. This proves equation (5.9) for almost every t.

From (5.9) we infer

$$R_\chi(t)L_\chi(t) = L_\chi(t)R_\chi(t) = h\int_K \chi(k)R_\chi(t)V_t(k)dk. \tag{5.10}$$

This proves

LEMMA 5.3. *The subspace $L_\chi(t)R_\chi(t)\mathfrak{H}_t$ of $R_\chi(t)\mathfrak{H}_t$ is for almost every t the sum of all those subspaces of $R_\chi(t)\mathfrak{H}_t$ which transform equivalently to the repersentation u of K under the operators $R_\chi(t)V_t(k)$, for $k\in K$. Here $R_\chi(t)V_t(k)$ is considered as an operator of the space $R_\chi(t)\mathfrak{H}_t$.*

Next we have

LEMMA 5.4. *The subalgebra $R_\chi(t)\mathbf{R}(t)R_\chi(t)$, considered as an algebra of operators of the space $R_\chi(t)\mathfrak{H}_t$ is for almost every t exactly the set of those*

bounded operators of the space $R_\chi(t)\mathfrak{H}_t$ which commute with (the restriction of) $R_\chi(t)V_t(g)$ (to the space $R_\chi(t)\mathfrak{H}_t$) for every $g \varepsilon G$.

Proof. According to (4.1) the ring $\boldsymbol{R}$ is the commuting algebra of $\boldsymbol{L}$ in the space $\mathfrak{L}_2(G)$. Hence if we apply Lemma 13 of [15] to our above central decomposition of $\mathfrak{L}_2(G)$, we may conclude that $\boldsymbol{R}(t)$ is the commuting algebra of $\boldsymbol{L}(t)$ for almost every t. By § 1 of [8] we know that after omission of one t-test of measure zero we have $V_t(g) \varepsilon \boldsymbol{L}(t)$ for all $g \varepsilon G$, and that the $V_t(g)$ generate $\boldsymbol{L}(t)$. But all this together with $R_\chi(t) \varepsilon \boldsymbol{R}(t)$ implies Lemma 5.4.

Let us now combine Lemmas 5.1, 5.2, 5.3 and 5.4. By Lemma 5.3 we know that $L_\chi(t)R_\chi(t)\mathfrak{H}_t$ is the sum of all those subspaces of $R_\chi(t)\mathfrak{H}_t$ which transform equivalently to the given repersentation $u(k)$ of K under the operators $V_t(k)$ for $k \varepsilon K$. Hence we have

$$(5.11) \quad \dim [L_\chi(t)R_\chi(t)\mathfrak{H}_t] = h \cdot [\text{multiplicity of } u \text{ in } R_\chi(t)V_t(K)].$$

Hence Lemmas 5.2 and 5.4 imply

$$(5.12) \quad \dim\{[R_\chi(t)V_t(g)]'\} = h \cdot [\text{multiplicity of } u \text{ in } R_\chi(t)V_t(K)],$$

where $[R_\chi(t)V_t(g)]'$ denotes the commuting algebra of the operators $R_\chi(t)V_t(g)$ in the space $R_\chi(t)\mathfrak{H}_t$.

Now let us consider the operators $U(g)$ of the induced representation $U = \operatorname{ind} u$ acting in the space $\mathfrak{H}$ as defined in **1**. The representation u of K being irreducible, we can apply Lemma 5.1 and conclude that the space $R_\chi\mathfrak{L}_2(G)$ can be identified with Kronecker product $\mathfrak{h} \times \mathfrak{H}$ such that $R_\chi L(g)$ becomes identified with $I_h \times U(g)$, where I_h denotes the identity matrix in the h-dimensional space $\mathfrak{h}$. If $\mathfrak{H} = \int_\oplus \mathfrak{U}_t$ is the central decomposition of $\mathfrak{H}$ under the operators $U(g)$ introduced in **1** (cf. equation (1.8)), then von Neumann's result on the essential uniqueness of the central decomposition (cf. loc. cit.) implies that the spaces obtained from the central decomposition of $R_\chi\mathfrak{L}_2(G)$ can be identified with the spaces $\mathfrak{h} \times \mathfrak{H}_t$ in such a manner that in particular each (measurable) operator-valued function in the one decomposition goes over into the corresponding operator-valued function in the other decomposition (neglecting of course sets of measure zero again). On the other hand we know from Lemma 3.1 that the spaces obtained from the central decomposition of $R_\chi\mathfrak{L}_2(G)$ may be identified for $t \varepsilon \boldsymbol{T}_1$ with the spaces $R_\chi(t)\mathfrak{H}_t$, where the $\mathfrak{H}_t$ are the component spaces of the central decomposition of $\mathfrak{L}_2(G)$ itself, so that corresponding operator-valued functions go again over

into each other. Hence we see that there is a one-one correspondence $t' \leftrightarrow t$ between almost all elements t' which occur in (1.8) and almost all those elements t which occur in (4.2) and are elements of $\boldsymbol{T}_1$. Here $\boldsymbol{T}_1$ is the set of those t for which $R_\chi(t) \neq 0(t)$. Moreover this one-one correspondence $t' \leftrightarrow t$ is such that for each corresponding pair t', t there exists a unitary operator $J(t', t)$ mapping $\mathfrak{H}_t$, onto $R_\chi(t)\mathfrak{H}_t$ in such a manner that

$$(5.13) \qquad R_\chi(t)V_t(g) = J(t', t)[I_h \times \mathfrak{V}_{t'}(g)]J(t', t)^{-1}.$$

Here the operators $\mathfrak{V}_{t'}(g)$ are the operators of the unitary representation of G acting in $\mathfrak{H}_{t'}$ as introduced in **1** (compare equations (1.8) and (1.9)).

From (5.13) we infer at once

$$(5.14) \qquad \dim\{[R_\chi(t)V_t(g)]'\} = h^2 \cdot \dim\{[\mathfrak{V}_{t'}(g)]'\}$$

for the dimensions of the commuting algebras of the operators $R_\chi(t)V_t(g)$ and $V_{t'}(g)$ respectively.

Also (5.13) obviously implies

$$(5.15) \qquad h^2 \cdot [\text{multiplicity of } u \text{ in } \mathfrak{V}_{t'}(K)]$$
$$= \text{multiplicity of } u \text{ in } R_\chi(t)V_t(K).$$

Hence combining (5.12) and (5.15) we get

$$(5.16) \qquad \dim\{[R_\chi(t)V_t(g)'\} = h^2 \cdot [\text{multiplicity of } u \text{ in } \mathfrak{V}_{t'}(K)]$$

for each corresponding pair t', t. Hence (5.14) and (5.16) together imply for almost every t'

$$(5.17) \qquad \text{multiplicity of } u \text{ in } \mathfrak{V}_{t'}(K) = \dim\{[\mathfrak{V}_{t'}(g)]'\}.$$

Since equation (5.17) is exactly the assertion of Theorem 1, i.e. equations (5.17) and (1.10) are identical, *the proof of Theorem* 1 *is herewith completed.*

Remark. In the course of this proof we have obtained somewhat more information than Theorem 1 asserts. For instance if we combine Lemmas 5.2, 5.3 and 5.4 we see that there is a *natural* linear one-one mapping between a dense linear subspace of the sum of those subspaces of $R_\chi(t)\mathfrak{H}_t$ which transform equivalently to $u(k)$ under the subgroup K and a dense linear subspace of the commuting algebra of the operators $R_\chi(t)V_t(g)$ in the space $R_\chi(t)\mathfrak{H}_t$.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] E. Cartan, "Sur la determination d'un system orthogonal complet dans un espace de Riemann symmetrique clos," *Rendiconti del Circolo Matematico di Palermo*, vol. 53 (1929), pp. 217-252.

[2] ———, "Les espaces Riemannienes symetriques," *Verhandlungen des Internationalen Mathematiker-Kongresses*, Zürich, 1932, vol. 1, pp. 152-161.

[3] ———, "Sur les domaines bornés homogènes de l'espace de n variables complexes," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 11 (1936), pp. 111-162.

[4] I. M. Gelfand, "Spherical functions in symmetric Riemann spaces," *Doklady Adademiia Nauk SSSR*, vol. 70 (1950), pp. 5-8.

[5] G. W. Mackey, "Imprimitivity for representations of locally compact groups. I," *Proceedings of the National Academy of Sciences*, vol. 35 (1949), pp. 537-545.

[6] ———, "Imprimitivité pour les representations des groupes localement compact II," *Comptes Rendus*, vol. 230 (1950), pp. 808-809, III., vol. 230 (1950), pp. 908-909.

[7] F. I. Mautner, "Unitary representations of locally compact groups I," *Annals of Mathematics*, vol. 51 (1950), pp. 1-25.

[8] ———, "Unitary representations of locally compact groups II," *ibid.*, vol. 52 (1950), pp. 528-556.

[9] ———, "The structure of the regular representation of certain discrete groups," *Duke Mathematical Journal*, vol. 17 (1950), pp. 437-441.

[10] ———, "Infinite dimensional irreducible representations of certain groups," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 582-584.

[11] ———, "On the decomposition of unitary representations of Lie groups," *Proceedings of the American Mathematical Society*, vol. 2 (1951), pp. 490-496.

[11a] ———, "Fourier analysis and symmetric spaces," *Proceedings of the National Academy of Sciences*, vol. 37 (1951), pp. 529-533.

[12] F. J. Murray and J. von Neumann, "Rings of operators," *Annals of Mathematics*, vol. 37 (1936), pp. 116-229.

[13] J. von Neumann, "On rings of operators III," *ibid.*, vol. 41 (1940), pp. 94-161.

[14] F. J. Murray and J. von Neumann, "On rings of operators IV," *Annals of Mathematics*, vol. 44 (1943), pp. 716-808.

[15] J. von Neumann, "On rings of operators. Reduction Theory," *ibid.*, vol. 50 (1949), pp. 401-485.

[16] ———, "Zur Algebra der Funktionaloperationen und Theorie der Normalen Operatoren," *Mathematische Annalen*, vol. 102 (1929), pp. 370-427.

[17] C. E. Rickart, "Banach algebras with an adjoint operation," *Annals of Mathematics*, vol. 47 (1946), pp. 528-549.

[18] C. L. Siegel, "Symplectic Geometry," *American Journal of Mathematics*, vol. 65 (1943), pp. 1-86.

[19] H. Weyl, "Harmonics on homogeneous manifolds," *Annals of Mathematics*, vol. 35 (1934), pp. 486-499.

[20] A. Wintner, "On the location of continuous spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 22-30.